

**A DYNAMIC HOLDING METHOD TO AVOID BUS
BUNCHING ON HIGH-FREQUENCY TRANSIT
ROUTES: FROM THEORY TO PRACTICE**

A Thesis Proposal
Presented to
The Academic Faculty

by

Simon J Berrebi

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Civil and Environmental Engineering

Georgia Institute of Technology
August 2017

Copyright © 2017 by Simon J Berrebi

A DYNAMIC HOLDING METHOD TO AVOID BUS BUNCHING ON HIGH-FREQUENCY TRANSIT ROUTES: FROM THEORY TO PRACTICE

Approved by:

Dr. Kari E Watkins, Advisor
School of Civil and Environmental
Engineering
Georgia Institute of Technology

Dr. Alan L Erera
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Dr. Jorge A Laval
School of Civil and Environmental
Engineering
Georgia Institute of Technology

Dr. Michael Meyer
Parsons Brinckerhoff

Dr. Michael O Rodgers
School of Civil and Environmental
Engineering
Georgia Institute of Technology

Date Approved: Saturday 15th July, 2017

To my grandmothers, Mamita, Mamie Annie, and Mamie Simone.

ACKNOWLEDGEMENTS

I have been immensely privileged to receive the mentorship of Dr. Kari Watkins in the last 5 years. Since the day I arrived at Georgia Tech with a confused urge to solve transit problems with theoretical math, she has advised me through each step of the research process, from theory to practice. Dr Watkins was a constant source of inspiration, enthusiasm, and vision. Her guidance went far beyond the confines of research, she has shaped me into who I am. For that, I will eternally be indebted to her.

I would like to thank my committee for their support. Dr. Jorge Laval gave me his expert advice on my research and on a thousand other irrelevant mathematical problem on which we inevitably ended side-tracking. Working with him was always a pleasure. Alan Erera gave me invaluable guidance in the early stages of this research. I thank Dr. Michael Rodgers with whom I worked in my first semester at Georgia Tech. I thank in particular Dr Michael Meyer, who has a unique way to think about transportation problem and to challenge my own thinking. His mentorship allowed me to reach further.

I wish to thank my collaborators across the Atlantic who helped me push this research into the real world. It was a pleasure to work with Dr. Etienne Hans and the ENTPE research team in Lyon, France. Merging my holding method with Etienne's prediction tool produced something that we could never have achieved separately. Sean Óg Crudden made the real world implementation of my holding method possible with his programming talent and his drive. Although we never met in person (he lives in Dublin) we have spent many sleepless nights together (especially him due to the time difference) building new functionalities and fixing software bugs. I am grateful

for the opportunity to work with such amazing researchers as Etienne and Sean.

I am grateful for the City of Atlanta, San Antonio VIA, and the Georgia Tech Parking and Transportation Department for allowing me to implement this research on their transit systems. At the City of Atlanta, Marwan Al-Mukhtar and Linda Lee agreed to run one additional streetcar for dozens of hours to test the holding method. I am thankful for the operators and dispatchers, in particular Tara Wright and Roy Waite, who worked overtime to implement the holding method. At San Antonio VIA, I would like to thank Manjiri Akalkotkar for her willingness to apply our research to a route used by 6000 people daily. I am also thankful for the VIA IT Department, Operations, Planning, and for the operators, dispatchers, and supervisors who implemented the method on the ground. At Georgia Tech, I would like to thank David Crites, David Williamson, and Gary Tavarez for supporting the implementation of the holding method. I am also grateful for the help from Dr. Russ Clark and Dr. John Bartholdi for their help getting the GPS trackers to work.

I am deeply appreciative for all the amazing people who have surrounded me with their camaraderie and wisdom over the last five years. I thank my best friend, Alice Grossman, for supporting me and exchanging critiques over thousands of double macchiatos (one wet one dry). I am grateful for having an extraordinary cohort of exceptional individuals who will soon run every aspect of transportation. I have shared countless moments of procrastination with Aditi Misra, Aaron Greenwood, Charlene Mingus, Amy Moore, Franklin Gbologah, Alper Akanser, Stephanie Yankson, Janille Smith-Colin, and many others. I hope to cross your paths more than just once a year. I also need to thank our program's graduate coordinator, Robert Simon, without whom I would have been thrown out of Georgia Tech long ago and on numerous counts.

I thank my friends and family who have supported me at every step of the way. The friends I have made in Atlanta are exemplary individuals who have given me

great inspiration. They know who they are and that I love them. I thank my parents and my sisters who sent me their love from overseas and visited Atlanta more than they thought. I missed them but never doubted their love. I miss my grandmothers, in particular Mamie Annie and Mamie Simone who left us recently. This dissertation is dedicated to them

Finally, thank you Hannah for holding my hand along this journey. Your thirst for adventure has filled my life with excitement, your tenderness has brightened my days, your love has carried me through. Thank you.

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	x
LIST OF FIGURES	xi
LIST OF SYMBOLS OR ABBREVIATIONS	xiii
SUMMARY	xiii
I INTRODUCTION	1
1.1 Background and Motivation	1
1.2 Research Approach	4
1.2.1 Study 1: Deriving Optimal Holding Method	4
1.2.2 Study 2: Comparing Methods by Simulation	4
1.2.3 Study 3: Live Implementation	6
II A REAL-TIME BUS DISPATCHING POLICY TO MINIMIZE PASSENGER WAIT ON A HIGH FREQUENCY ROUTE	8
2.1 Abstract	8
2.2 Introduction	8
2.3 Literature Review	11
2.4 Formulation	12
2.4.1 Structural properties of an optimal policy	16
2.4.2 Proposed control method	20
2.4.3 Numerical example	23
2.5 Simulation	24
2.5.1 Instability	26
2.5.2 Perturbation	28
2.6 Conclusion	30

III	COMPARING BUS HOLDING METHODS WITH AND WITHOUT REAL-TIME PREDICTIONS	32
3.1	Abstract	32
3.2	Introduction	32
3.3	Holding methods in the literature	35
3.3.1	Naive methods	36
3.3.2	Partial holding methods	37
3.3.3	Prediction-based holding methods	38
3.4	Case study	40
3.4.1	The route	40
3.4.2	Prediction	42
3.4.3	Performance indicators	46
3.5	Cross-comparison	47
3.6	Sensitivity	50
3.6.1	Parameterization	50
3.6.2	Multiple Control Points	53
3.7	Prediction	57
3.8	Conclusion	60
IV	IMPLEMENTING A DYNAMIC HOLDING METHOD ON THREE HIGH-FREQUENCY TRANSIT ROUTES	63
4.1	Abstract	63
4.2	Introduction	63
4.3	Literature Review	67
4.4	Deriving Deterministic Method	68
4.5	Case Study	71
4.5.1	DynamicTime	71
4.5.2	Transit Systems	72
4.6	Results	76
4.6.1	Atlanta Streetcar	76

4.6.2	San Antonio VIA	79
4.6.3	Georgia Tech	87
4.7	Lessons Learned	89
4.7.1	Location Data	89
4.7.2	Prediction	92
4.7.3	Human element	93
4.7.4	Surrounding environment	94
4.8	Conclusion	95
V	CONCLUSION	98
5.1	Contributions	98
5.1.1	A Real-Time Dispatching Policy	98
5.1.2	Simulation of Holding Methods	99
5.1.3	Implementation on Three High-Frequency Routes	100
5.2	Concluding Remarks and Future Research	100
5.2.1	Application to broader transit problems	102
	APPENDIX A	105
	BIBLIOGRAPHY	109

LIST OF TABLES

1	Summary of variable definitions.	13
2	Summary of variable definitions.	37
3	Holding methods with their data requirements and recommended holding times.	39
4	Particle generation	44

LIST OF FIGURES

1	Arrivals, departures, and holds from the control point.	14
2	Bus arrival and optimal departure process at a control point	21
3	Optimal hold h_1 as a function of correlation between A_2 and A_3	24
4	Mean passenger waiting time versus level of systemic instability	27
5	Mean headway versus level of systemic instability	28
6	Headway coefficient of variation versus level of systemic instability	28
7	Mean passenger waiting time versus number of dispatches	29
8	Map of TriMet Route 72 TriMet (2016)	41
9	(a) Generation of particles to simulate bus trajectories. (b) Histograms of simulated bus arrival times treated as probability distributions.	45
10	Performance of holding methods in terms of (a) squared headway coefficient of variation, (CV^2) and (b) mean holding time past boarding and alighting (lost time).	48
11	(a) CV^2 as a function of α , (b) mean lost time as a function of α and (c) CV^2 as a function of mean lost time at station 49.	52
12	Mean CV^2 and total lost time applied at 1,2,4, and 8 control points for all methods and at 16 and 32 control points for the Naive Schedule and the methods in Xuan et al. (2011).	55
13	Sensitivity of (a) CV^2 and (b) Mean holding time to the maximal error in prediction	59
14	Screen capture of the DynamicTime dashboard	72
15	Map of the Atlanta Streetcar route adapted from Atlanta (2016)	74
16	Map of VIA Route 100 (VIA, 2017)	75
17	Map of Georgia Tech Red Stinger Route (Parking and Services, 2017)	76
18	Headways at departure from the control point (black), actual holding times (gold), and holding times recommended by the DynamicTime application (blue) as a function of time.	78
19	Histogram of headways of at departure from the control point as dispatched by the schedule (white) and by the proposed method (black).	80

20	Histogram the actual holding times in seconds for the schedule method (white) and the proposed method (gold), and the recommended holding times by the proposed method (blue).	81
21	Headways and holding times from the proposed method at the northern (top) and southern (bottom) control points on Tuesday, May 9. . . .	82
22	Headway Coefficient of Variation (CV^2) throughout the route between 2:30 and 5:30 under the schedule (January 3, 17, and 31) and under the proposed method (May 9)	84
23	Headways and holding times from the proposed method at the northern (top) and southern (bottom) control points on Friday, May 12.	86
24	Headway Coefficient of Variation (CV^2) throughout the route between 2:30 and 5:30 under the schedule (January 12 and 20) and under the proposed method (May 12)	88
25	Headway Coefficient of Variation (CV^2) throughout the route between 2:30 and 5:30 under the schedule (January 12 and 20) and under the proposed method (May 12)	90

SUMMARY

On frequent transit routes, there is a tendency for vehicles to bunch together, causing undue passenger waiting time. To avoid bus bunching, transit agencies strive to maintain even headways by holding vehicles at control points. The increasing availability and accuracy of real-time information enables transit agencies to apply dynamic holding methods that can stabilize their high-frequency routes based on the current operating conditions. This dissertation explores how real-time information can be used to stabilize high-frequency routes, from theory to practice.

In the first part of this dissertation, a closed-form method to avoid bus bunching is derived by backward induction as a non-Markovian stochastic decision process. The method can minimize expected headway variance without using buffer time: the rate of vehicle dispatch is equal to the rate of arriving vehicles. The decision process is based on a set of arrival time probability distributions that is updated at each decision epoch. The method is therefore able to minimize the waiting time of passengers arriving at stops according to a Poisson Process. The closed-form method is verified by simulation

In the second part of this dissertation, the method derived previously is compared to holding methods used in practice and recommended in the literature. The methods are tested on a simulation of Tri-Met Route 72 in Portland, OR, using historical data. A series of sensitivity analyses are conducted to evaluate the impact of parameter choice, control point selection, and prediction accuracy on the performance of each method. The methods based on predictions, and in particular the proposed method, were found to yield the best compromise between holding time and headway stability. We found that prediction errors had a marginal impact on

the performance of these methods until they reached a breaking point, beyond which they had a disproportionate effect.

In the third part of this dissertation, the proposed holding method is implemented in three high-frequency transit routes: the Atlanta Streetcar, the VIA Route 100 in San Antonio TX, and the Georgia Tech Red Stinger Route. The performance of the proposed method is compared to the schedule that is currently used. The different institutional frameworks of the agencies that collaborated on this project and their impact on implementation are compared. In addition, the level of adoption by supervisors, dispatchers and operators and their effect on compliance are analyzed. Finally, the practical lessons learned from implementing a real-time dispatching method on high-frequency routes are discussed.

Large parts of this thesis proposal were extracted from the following papers:

- Berrebi, S., Watkins, K., Laval, J., 2015. A Real-Time Bus Dispatching Policy to Minimize Passenger Wait on a High Frequency Route. Transportation Research Part B: Methodological
- Berrebi, S., Hans, E., Chiabaut, N., Laval, J., 2016. Leclercq, L., Watkins, K., Comparing Bus Holding Methods using Real-Time Predictions. Transportation Research Part C: Technological. (Forthcoming)
- Berrebi, S., Crudden, S., Ó., Watkins, K., 2017. Implementing Real-Time Holding Method: From Theory to Practice. Working paper

CHAPTER I

INTRODUCTION

1.1 Background and Motivation

High-capacity transit generates economic growth, promotes healthy lifestyle, and provides access to opportunities while minimizing the negative externalities of transportation. The high person-throughput of these routes make the most effective use of limited right-of-way and energy resources. According to the American Public Transportation Association, transit saves the equivalent of 4.2 billion gallons of gasoline and 37 million metric tones of carbon dioxide annually. To yield the benefits of shared mobility, transit agencies need to maximize ridership on their high-capacity routes. Transit agencies, however, are currently facing mounting competition from demand-responsive services in low-occupancy vehicles. These emerging modes are attractive to transit users because they are available on-demand. In order to remain competitive, transit agencies need to set a new standard for reliability and give users the freedom to make spontaneous travel plans and to change them at the last minute.

The current paradigm for transit reliability is based on the agency's ability to meet a schedule. The schedule, however, obliges users to plan around a set time-frame. The schedule adherence, therefore, does not correspond to users perception of reliability on these routes. Instead, perceived reliability and mode choice are driven by at-stop waiting time, which is weighted by passengers twice as much as in-vehicle time according to the Transit Capacity and Quality of Service Manual (KFH, 2013).

Users tend to disregard the schedule altogether and arrive randomly on routes with headways shorter than 12 minutes (Fan and Machemehl, 2009). For passengers who arrive at stop according to a Poisson process, waiting time is proportional to

both the average headway and the headway variance. When buses bunch together, passengers face large gaps in service and disproportional waiting times. Because more passengers arrive during long headways than during short ones, and because these passengers have to wait especially long, passenger wait is proportional to both mean headway and headway variance (Newell and Potts, 1964). In order to provide reliable service and compete with the full range of other modes, transit agencies need to maintain short and stable headways.

On high frequency routes, there is a natural tendency for buses to *bunch* together. When a bus is traveling with a long headway¹, it has to pick up and drop off a relatively greater number of passengers, which slows down the bus even more. As the lagging bus becomes crowded, the following buses only have a few passengers to serve, which takes less time. Eventually the lead bus gets caught by one or several following buses and they start traveling as a platoon. Bus bunching is the product of unstable dynamics that cause delays to grow (Hickman, 2001). Even a small perturbation such as a traffic signal or a passenger paying in cash can destabilize the route and lead to bus bunching (Kittelson, 2003).

When scheduling frequent transit service, agencies strive to allocate their limited capacity to the demand and prevent large gaps in service from forming. Scheduling reliable service is difficult because bus running times constantly fluctuate due to traffic, passenger demand, weather, etc. Based on historical data, transit agencies include buffer time in their operations (Furth et al., 2006). Scheduled running time is usually set as the 80th to the 95th percentile of observed running time (Levinson, 1991; Boyle et al., 2009). Most of the time, however, vehicles finish their assignments early and wait in layover at the terminal station when they could be running the route. Buffer time costs precious resources to the agencies that could otherwise be used to

¹In this text, the term *headway* will be used as the time since the leading vehicle passed the current location. Later, we will distinguish between the headway (or forward headway) and the backward headway, which is the time until the following vehicle will reach the current location.

serve passengers. On the other hand, when operating conditions are congested, the system tends to destabilize quickly as demand accumulates, and the buffer capacity is seldom able to bring back the system into equilibrium. Therefore, the schedule does not correspond to users' perception of reliability and is unable to maintain efficient operations on high frequency routes.

The increasing accuracy and availability of real-time information allows transit agencies to replace the static decision making of schedules by real-time holding methods (Development, 2013). Real-time vehicle location data make it possible to predict the evolution of the route and to identify the underlying points of equilibrium. Based on this information, vehicles can be held at control points along the route according to auto-correcting operational rules that can stabilize the system. These methods can diffuse large gaps in service based on current operating conditions, without requiring substantial buffer time.

There are two main families of real-time dispatching methods in the literature. The first consists in simulating the future states of the system and to minimize a weighted function of passenger wait on a rolling horizon. Simulation-based methods can take a holistic approach to solving the bus bunching problem by considering the longer-term effects of decisions. These methods, however, rely on black-box models to predict the evolution of inherently chaotic systems. They may prescribe a course of action without considering the small disruptions that may ultimately destabilize the route.

The second approach consists in formulating closed-form solutions to the bus bunching problem based on vehicles headways and predicted arrival times. Closed-form methods are intuitive and allow a greater insight into the bus bunching problem. These methods, however, only ever consider the vehicles immediately adjacent to the control point. Since bus bunching is a global problem, it cannot be solved efficiently by simply taking a local perspective. There lacks a closed form method that can solve

the bus bunching problem while considering every vehicle on the route. This dissertation explores closed-form self-correcting mechanisms that use real-time information to bring the system back into the time-dependent equilibrium.

1.2 Research Approach

1.2.1 Study 1: Deriving Optimal Holding Method

The waiting time of randomly arriving passengers on a high-frequency route is directly proportional to the sum of squared headways (Osuna and Newell, 1972). The solution to the deterministic least square problem was invented by Gauss in *Theoria motus corporum coelestium in sectionibus conicis solem ambientum* (Gauss, 1809). The bus dispatching problem, however is stochastic: predicted arrivals have probability distributions that shift and slim as the vehicle gets closer at each decision epoch (i.e. each time a vehicle arrives at a control point and a holding decision needs to be made). The problem should therefore be addressed sequentially as a stochastic decision problem. This decision process is difficult to solve in closed form because it can be summarized by its current state only when considering all possible combinations of headways at previous epochs.

In Chapter 2, the bus dispatching problem is solved as a stochastic decision problem. The optimal holding policy (i.e. systematic decision rule) is derived by backward induction. A discrete random variable is introduced to treat the edge cases due to the constraint that holds must always be equal to or greater than zero. A method is proposed to approximate the discrete variable. Finally the method is compared to the Naive Headway and the method from Bartholdi and Eisenstein (2011) in a simulation.

1.2.2 Study 2: Comparing Methods by Simulation

Holding buses at control points can help to mitigate bus bunching, and even prevent it from happening if applied correctly; but holding vehicles also reduces their overall

operating speed. If holds are applied at a mid-route control point, then holding time delays passengers who are already boarded. There is a trade-off between stabilizing headways and maintaining high operating speed (Furth et al., 2006; Furth and Wilson, 1981; Cats et al., 2011). Transit agencies value these conflicting objectives differently depending on arrival time distributions, travel distances, connection modes, route geometry, etc. Accordingly, a holding method that stabilizes the route but requires long holding times may be better suited for certain types of routes than for others and *vice versa*. In addition, parametrization, density of control points, and quality of the data all impact a method’s ability to stabilize the route and its required holding time. Transit agencies need the tools to identify the most suitable holding method for their routes.

In Chapter 3, we investigate how closed-form holding methods used in practice and recommended in the literature compromise between headway stability and holding time. To this end, each method is simulated using historical data from Tri-Met Route 72 in Portland, Oregon. A series of sensitivity analyses are presented to compare the relative advantages and disadvantages of each method and the implications of parameter choice, control point selection, and prediction accuracy. Follows a discussion on the relative adequacy of each method based on the type of application.

We expanded the prediction tool developed in Hans et al. (2015) to reproduce the predictions in a realistic setting. The prediction method is a particle filter combined with an event-based mesoscopic model, which can generate the arrival time distributions required by the method presented in Chapter 2. This work was conducted in collaboration with Dr. Étienne Hans, Dr. Nicolas Chiabaut, and Dr. Ludovic Leclercq. Dr. Étienne Hans developed the code to expand his prediction tool into a simulation environment. My main contributions were to create the specifications for the case study, the holding methods and the sensitivity analyses. I wrote code to increase the density of analysis on each figure and wrote the paper. Finally, I

analyzed the data from the simulation, and discussed their implication.

1.2.3 Study 3: Live Implementation

There is a need to evaluate the performance the proposed method in a real setting. The simulation experiments conducted thus far in Berrebi et al. (2015) and Berrebi et al. (2016) showed that the proposed holding method can dispatch vehicles with more stable headways than methods used in practice and recommended in the literature with the same mean holding time. These simulation experiments, however are limited to theoretical interpretations of a transit route. They cannot consider all the small perturbations that can accumulate and lead to the destabilization of the route. Therefore, simulation experiments may not fully represent the compounded effects of holding methods on transit operations and reciprocally.

In the third part of this dissertation, we implement the holding method developed in Chapter 2 on three high-frequency transit routes: the Atlanta Streetcar, the VIA Primo 100 Route, and the Georgia Tech Red Stinger Route. These three routes offer widely different test cases. The first is a streetcar route running in the heart of Downtown Atlanta, the second is a Bus Rapid Transit route connecting San Antonio suburbs to downtown through Highway I-10, and the third serves a student population, which surges the system before and after class. Each route runs in mixed traffic on a schedule that is unavailable to the public. The study discusses the lessons learned from implementing a real-time dispatching method in high-frequency transit systems.

DynamicTime, the software used for the proposed method was derived from the open source application TransiTime. TransiTime is a collaborative project that generates predicted vehicle arrivals based on Automatically Vehicle Location (AVL) data. The research is the product of a collaboration with open-source developer Sean Óg Crudden. Sean Crudden expanded TransiTime to support frequency-based service.

He also wrote all the code to create DynamicTime. My contributions were to create the specifications for DynamicTime, and to inform the software design with the feedback I received from Atlanta Streetcar dispatchers. I tested the DynamicTime platform and fixed the bugs with Sean Crudden. I assisted the Atlanta Streetcar and VIA dispatchers. On the Georgia Tech Red Route, I communicated holding instructions to operators personally, with the help of Undergraduate Research Assistant Reid Passmore. Finally, I performed all the analysis and drafted the paper.

CHAPTER II

A REAL-TIME BUS DISPATCHING POLICY TO MINIMIZE PASSENGER WAIT ON A HIGH FREQUENCY ROUTE

2.1 *Abstract*

One of the greatest problems facing transit agencies that operate high-frequency routes is maintaining stable headways and avoiding bus bunching. In this work, a real-time holding mechanism is proposed to dispatch buses on a loop-shaped route using real-time information. Holds are applied at one or several control points to minimize passenger waiting time while maintaining the highest possible frequency, i.e. using no buffer time. The bus dispatching problem is formulated as a stochastic decision process. The optimality equations are derived and the optimal holding policy is found by backward induction. A control method that requires much less information and that closely approximates the optimal dispatching policy is found. A simulation assuming stochastic operating conditions and unstable headway dynamics is performed to assess the expected average waiting time of passengers at stations. The proposed control strategy is found to provide lower passenger waiting time and better resiliency than methods used in practice and recommended in the literature.

2.2 *Introduction*

High frequency transit services, including rail, Bus Rapid Transit, and enhanced bus¹, give users the freedom to decide when and how to travel; it allows them to make last minute plans and to change their route spontaneously (Walker, 2011). When

¹For simplicity, transit vehicles will be referred to as buses for the remainder of this text.

buses run every 11 minutes or less, transit riders tend to arrive at stations without consulting a schedule, and to wait for the next passing vehicle (Fan and Machemehl, 2009). Because passengers have proportionally more chance of arriving during a long headway than during a short one, passenger waiting time is proportional to the sum of squared headways (Osuna and Newell, 1972). Headway regularity also defines the reliability of a bus route, and the amount of trust that passengers can give it (Program, 2013; Furth and Muller, 2007). To avoid undue passenger waiting time, and to promote reliability, transit agencies strive to maintain headways as even as possible.

Transit routes are naturally unstable systems; headway variance tends to increase along a route, causing buses to bunch together (Strathman et al., 2002; Boyle et al., 2009; Hickman, 2001; Daganzo, 2009; Program, 2013). When a bus arrives at a station with a long headway, a large number of passengers must board (and alight), causing further delay. As the delay of a lagging bus tends to grow, the following bus can operate at a higher speed because it does not need to board (and alight) as many passengers. Eventually, the following bus tends to catch up to the lagging bus; we call this bus bunching. Dispatching buses from a control point with disproportionate headways can trigger bus bunching from the onset (Levinson, 1991).

It is difficult for transit agencies to schedule both reliable and efficient operations because bus running time can be quite variable due to fluctuating operating conditions such as traffic, passenger demand, weather, etc. A bus that has accumulated delays from its last assignment and arrived late at control point may have to be dispatched with a long headway. This happens when schedulers plan tight operations with short scheduled running times (Boyle et al., 2009). To avoid the formation of *big gaps*, transit agencies include buffer time in their operations. Scheduled running time is usually set as the 80th to the 95th percentile of observed running time (Furth et al., 2006; Levinson, 1991; Boyle et al., 2009). The majority of buses finish their

assignments faster than the scheduled running time. These buses must wait in layover at the terminal station instead of running another route. In addition, the running times of subsequent vehicles depend on current operating conditions. In high levels of congestion, all buses are slowed down. When this occurs, the buffer time becomes insufficient to avoid buses starting their route with irregular headways.

Since the deactivation of selective availability of GPS in 2000, Automatic Vehicle Location (AVL) technologies have been increasingly available and reliable. In 2013, the American Public Transportation Association surveyed seventy-five transit agencies and more than 70% reported that they could track buses in real-time. Of the agencies that did not have access to these technologies, 92% were interested in adopting them (Development, 2013). With real-time information, transit agencies can now predict travel time with high accuracy (Chien et al., 2002; Cathey and Dailley, 2003; Jeong and Rilett, 2004b; Shalaby and Farhan, 2004; Gurmu and Fan, 2014; Ding, 2000). Transit agencies can replace the static decision making of schedules by informed operational decisions to dispatch buses with low headway variance without requiring substantial buffer time.

In this paper, the bus dispatching problem is addressed as a stochastic decision problem. The objective is to minimize passenger waiting time, while maintaining the maximal frequency at current operating conditions, i.e. to minimize the sum of squared headways. The following section reviews methods to mitigate bus bunching in the literature. In Section 3, structural properties of an optimal bus dispatching policy are derived, a streamlined control method is proposed, and a numerical example is presented. In Section 4, a bus route is simulated and the proposed control strategy is compared to methods used in practice, and recommended in the literature.

2.3 Literature Review

In the literature, multiple on-route methods are used to apply control at intermediate locations to stabilize headways (Barnett, 1974; Eberlein et al., 2001; Zolfaghari et al., 2004; Oort et al., 2009). These methods minimize passenger waiting time assuming that perfect predictions are available and that buses depart the terminal station on time, but tend to bunch together due to unstable headway dynamics. Daganzo et al. solved a similar problem, considering random travel times (Daganzo, 2009). They included an *instability parameter* that accounted for the headway dynamics. On headway-based routes, Van Oort et al. and Xuan et al. sought to reduce deviations from the *planned headway* by holding vehicles at intermediate control points (Oort et al., 2009; Xuan et al., 2011). Hickman modeled headway dynamics as a Markov Chain and minimized the mean passenger waiting time by considering one headway at a time (Hickman, 2001). Delgado et al. minimized total waiting time, considering boarded and waiting at-stop passengers in static operating conditions. These on-route control methods hold vehicles one at a time, without considering future decisions. They stabilize headways locally by using the buffer time built into planned operations. The amount of buffer time is determined off-line, so it may be insufficient sometimes and excessive at other times.

Osuna and Newell studied the amount of buffer time required to minimize passenger waiting time using stationary cycle time probability distributions (Osuna and Newell, 1972; Newell, 1974). They formulated the bus dispatching problem sequentially as an infinite horizon Markov Decision Process (MDP) and found approximately optimal policies for up to two vehicles. In both papers, the problem became intractable when several vehicles and stations were introduced. Zhao et al. developed a queuing model to optimize buffer time in schedules on a high frequency route (Zhao et al., 2006). The method does not use real-time information, and assumes stationary operating conditions.

Bartholdi and Eisenstein addressed for the first time the problem of bus dispatching by blending headways using real-time information (Bartholdi and Eisenstein, 2011). They found a powerful heuristic to send buses with low mean headway and headway variance. When a bus arrives at a control point, it is held for a fraction of the predicted time until the next arrival, or until its headway reaches the *planned headway*, whichever is greater. The method averages adjacent headways to reduce their difference, and to split *big gaps* in half. The *planned headway* is needed to prevent bunched vehicles, which arrive in close succession, from leaving with too short of headways.

The real-time control methods in the literature address the dispatching problem locally, without considering the arrival time of subsequent vehicles. The delay of a bus is only considered into the hold of its predecessor. In this paper, we present a global approach to the bus dispatching problem such that *big gaps* get absorbed by several or all preceding buses. The objective is to minimize expected headway variance of every bus on the route at one or several control points, while maintaining maximal frequency.

2.4 *Formulation*

In this section, the bus dispatching problem is addressed as a finite-horizon stochastic decision process. The bus arrival and dispatching process is explained and the notation is introduced. Passenger waiting time is expressed as a function of the sum of squared headways. A dispatching policy to minimize the sum of squared headways is derived by backward induction in terms of expected arrival times. A streamlined approximation to the optimal policy is found and compared to the optimal policy in a numerical example. The following table is a glossary of variables used in this work.

In this paper, the framework of the bus dispatching problem is simplified to represent its most essential elements: buses run continually on a loop-shaped route with

Table 1: Summary of variable definitions.

Variable	Definition
n	Number of buses considered
A_i	Arrival time of bus i (Random Variable)
a_i	Arrival time of bus i (Realization)
D_i	Departure time of bus i (Random Variable)
d_i	Departure time of bus i (Realization)
h_i	Hold imposed on bus i
u_i^π	Expected sum of squared headways from epoch i
I_i^j	Indicator variable equal to 1 when the first zero hold after i is j .

one or several control points. When a bus arrives at a control point, it can hold for some time or start its next cycle immediately. The purpose of this paper is to find a holding policy that dispatches buses with stable headways, while maintaining the frequency as high as possible for the next n dispatches. The horizon n is the number of buses that are currently running between control points. If there is a single control point, then n is the number of buses on the route.

The first bus to arrive at the control point is called bus 1, and we index subsequent buses in increasing order.² For example, the bus that follows bus 1, i.e. that arrives at the control point right after bus 1, is bus 2, and so on. We say that bus 1 arrives at the control point at time a_1 , that it holds for h_1 time units, and that it departs the control point at time d_1 . The arrival times of following buses $\{2, \dots, n\}$ are random. We denote them $\{A_2, \dots, A_n\}$ and their outcomes are $\{a_2, \dots, a_n\}$. Once bus i arrives at the control point, it holds for a time h_i , which must be non-negative, because buses cannot depart the control point before they arrive. After the hold, bus i is re-dispatched onto the route at time D_i . The departure time is random because it is a function of A_i as shown below:

$$D_i = A_i + h_i \quad (1)$$

²To keep the notation as simple as possible, we consider that a bus has *arrived* at the control point when the operator has finished his or her break. It then becomes available to either hold or to start a new cycle.

The outcome of the random departure time of bus i is denoted d_i . Its equation is given below:

$$d_i = a_i + h_i \quad (2)$$

In Figure 1, the probability distributions of arrival (blue) and departure (green) times from the control point are displayed for n buses running on the route. A decision to hold bus 1 needs to be made at time a_1 . At that time, D_0 , A_1 , and D_1 are known exactly with probability one; the two former have already happened while the latter is to be determined immediately. Although the arrival times of subsequent buses are co-dependent, their probability distributions are shown as bell-shaped curves. Note that the arrival time of a bus is independent of preceding holds because the remainder of its route is upstream of the control point.

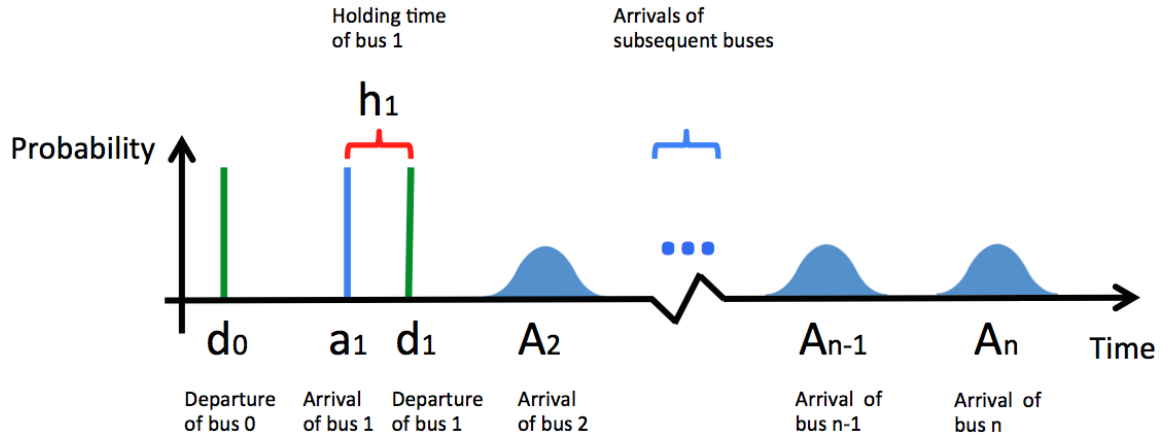


Figure 1: Arrivals, departures, and holds from the control point.

The headway of bus i is the time between its departure from the control point and the departure of the preceding bus, i.e. $D_i - D_{i-1}$. We know from Osuana and Newell (1972); Newell (1974) that the average waiting time for a passenger randomly arriving between d_0 and D_n is equal to the sum of squared headways divided by the sum of headways during which he or she arrives as shown in Equation (3).

$$\text{Average passenger wait} = \frac{\sum_{i=1}^n [D_i - D_{i-1}]^2}{2 \sum_{i=1}^n [D_i - D_{i-1}]} = \frac{\sum_{i=1}^n [D_i - D_{i-1}]^2}{2(A_n + h_n - d_0)} \quad (3)$$

A solution to the bus dispatching problem must consider sequentially the holds that will be imposed on upstream buses once they reach the control point. Those decisions will be supported by updated information and probability distributions. Taking into consideration the denominator of average passenger wait in Equation (3), $2(A_n + h_n - d_0)$, the term h_n can be interpreted as buffer time. On schedule-based and headway-based routes, buffer time is used to reduce the headway variance, however, buffer time also increases the mean headway. We seek to minimize the sum of squared headways, $\sum_{i=1}^n [D_i - D_{i-1}]^2$ assuming that h_n will be null, i.e. using no buffer time.

We say that the expected total cost criterion, denoted u_1^π , knowing the holding and arrival time history, is the sum of squared headways if holds are determined according to the systematic dispatching policy π going forward in time. This problem is a stochastic least square problem. Its criterion can be expressed iteratively as in (4) and expanded as in (5) by the Law of Iterated Expectations:

$$u_i^\pi(h_{i-1}^o, A_2, \dots, A_i) = (a_i - d_{i-1} + h_i^\pi)^2 + E[u_{i+1}^\pi(h_i^\pi, A_2, \dots, A_i)] \quad (4)$$

$$= (a_i - d_{i-1} + h_i^\pi)^2 + \sum_{j=i+1}^n E[(A_j - A_{j-1} + h_j^\pi - h_{j-1}^\pi)^2 | A_2, \dots, A_i] \quad (5)$$

Finding a dispatching policy π^* that minimizes u_i^π is equivalent to minimizing the variance of headways from the control point. Hickman derived the headway dynamics that lead to bus bunching and showed that headway variance tends to increase along a route (Hickman, 2001). Dispatching buses with stable headways is the best way to prevent bus bunching from the onset.

2.4.1 Structural properties of an optimal policy

In the remainder of this section, we derive structural properties of an optimal policy. Bellman's principle of optimality is applied to separate the sequential dispatching problem into $n - 1$ overlapping sub-problems. The i^{th} sub-problem consists in finding the optimal policy starting at decision epoch i as a function of $\{h_{i-1}, A_2, \dots, A_i\}$. Lemma 2.4.1 says that if a sub-policy starting at epoch i is optimal, then it is part of the optimal policy starting from the beginning. The proof is consistent with that of Theorem 4.3.2 in Markov Decision Processes, and can be found in Appendix A (Puterman, 2009).

Lemma 2.4.1. *Given h_{i-1}^o and $\{A_2, \dots, A_i\}$, for any i , the action h_i^* is defined as follows:*

$$h_i^* = \arg \min_{h_i} \left[(a_i - d_{i-1} + h_i)^2 + \sum_{j=i+1}^n E[(A_j - A_{j-1} + h_j^* - h_{j-1}^*)^2 | A_2, \dots, A_i] \right] \quad (6)$$

At each decision epoch, $\{h_i^, \dots, h_{n-1}^*\}$ minimizes $u_i^\pi(h_{i-1}^o, A_2, \dots, A_i)$.*

Any sub-policy that does not include h_1 should be viewed as part of a sub-problem that needs to be solved in order to determine the optimal holding policy and immediate action. Lemma 2.4.2. demonstrates that $u_i^\pi(h_{i-1}^o, A_2, \dots, A_i)$ is convex in $\{h_i, \dots, h_{n-1}\}$. This result will be used to derive the optimal dispatching sub-policies. The proof of Lemma 2.4.2. can be found in Appendix A.

Lemma 2.4.2. *The total expected cost criterion is convex in the action space.*

In this work, vehicles can only be held in the positive time direction, so the feasible region of a policy π is \mathbb{R}^+ , and is therefore convex. Any inflexion point in the total cost criterion inside the feasible region is a global minimum (Winston and Goldberg, 2004). If there exists no inflexion point inside the feasible region, then the closest boundary point to an inflexion point is a global minimum, i.e. no hold. In the following corollary, h_i^* is derived as a function of the future departure headways and their derivative with respect to the decision i .

Corollary 2.4.3. *Given h_{i-1}^o and $\{A_2, \dots, A_i\}$, the optimal action h_i^* as defined in Lemma 2.4.1 is equal to:*

$$h_i^* = \left[-a_i + d_{i-1} - \sum_{j=i+1}^n E \left[\frac{d[-h_{j-1}^* + h_j^*]}{dh_i} [A_j - A_{j-1} + h_j^* - h_{j-1}^*] | A_2, \dots, A_i \right] \right]^+ \quad (7)$$

Proof. The derivative of $u_i^\pi(h_{i-1}^o, A_2, \dots, A_i)$ with respect to h_i is:

$$2(a_i - d_{i-1}^o + h_i) + 2 \sum_{j=i+1}^n E \left[\frac{d[-h_{j-1}^* + h_j^*]}{dh_i} [A_j - A_{j-1} + h_j^* - h_{j-1}^*] | A_2, \dots, A_i \right] \quad (8)$$

If there exists an h_i such that (8) equals zero, then it is h_i^* . If there exists no such h_i , then by convexity of $u_i^\pi(h_i, A_2, \dots, A_i)$, $h_i^* = 0$. \square

The main difficulty in evaluating (7) comes from the non-linearity of the term $\frac{dh_j^*}{dh_i^*}$ because the effects of an immediate decision on the cost criterion at epoch i is subject to the hold recommended by an optimal policy at that time. If holds could be negative, then $\frac{dh_j^*}{dh_i^*}$ would be $\frac{n-j}{n-i}$ and $h_i = \frac{E[A_n - d_{i-1}]}{n-i+1}$. However, due to the constraint $h_j \geq 0$, h_j^* is not continuously differentiable at 0 with respect to h_i . Whether or not h_j^* will be positive, for a given action h_{i-1}^o and history $\{A_2, \dots, A_i\}$, depends on the outcome of the random variables $\{A_{i+1}, \dots, A_j\}$. The expectation in (7) is evaluated with respect to every possible combination of arrival times, including all those that would yield a null derivative and all those that would not. A classical Dynamic Programming (DP) approach to this problem would require discretizing time and storing in memory every possible combination of arrival times and actions, along with their associated probabilities. For this reason, we investigate further structural properties of an optimal policy. So far, no assumption was made on the joint probability distributions of the arrival times $f(a_2, \dots, a_n)$. To maintain the generality of the dispatching problem, let us define the following indicator random variable:

Definition.

$$I_j^i = \begin{cases} 1 & \text{if } h_q^* > 0 \ \forall q \in \{i+1, \dots, j-1\} \text{ and } h_j^* = 0 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

The term I_j^i corresponds to the event that bus j will be the first not to be held after bus i under an optimal bus dispatching policy. The random variable I_j^i is determined by the outcome of the inter-arrival times $\{A_{i+1}, \dots, A_j\}$. At the i^{th} decision epoch, $E[I_j^i]$ is the probability that the optimal policy will hold every bus between i and j , but will re-dispatch bus j as soon as it arrives. Note that $\sum_{j=i+1}^n E[I_j^i] = 1$ because n^{th} hold is null by definition, and that $E[I_{j_1}^i \cap I_{j_2}^i] = 0$ because the first active hold cannot happen at two different times. Taking the probabilities associated with the indicator variables, we get the following equations:

$$E[I_j^i] = P(h_{i+1}^* > 0, \dots, h_{j-1}^* > 0, h_j^* = 0 | h_i > 0) \quad \forall j > i \quad (10)$$

With this information, the agent considers how many headways will absorb the *big gap* preceding bus i . The following theorem is central to the evaluation of an optimal policy for the bus dispatching problem. The indicator random variables I_j^i are used to separate (7) in piecewise linear terms. In Theorem 2.4.4, the optimal holding policy is found in terms of the expected arrival time of future buses and the probability of holding them under the optimal policy.

Theorem 2.4.4. *At each decision epoch i , given h_{i-1} and $\{A_2, \dots, A_i\}$, the optimal holding policy is:*

$$h_i^* = \left[\frac{\sum_{q=i+1}^n E[I_q^i] \frac{E[A_q - A_i | A_2, \dots, A_i, I_q^i]}{q-i} + d_{i-1} - a_i}{1 + \sum_{q=i+1}^n \frac{E[I_q^i]}{(q-i)}} \right]^+ \quad (11)$$

Proof. By definition $h_n = 0$. Therefore $E[I_n^{n-1}] = 1$. The rhs of (11) for $i = n-1$ is:

$$\left[\frac{E[A_n | A_2, \dots, A_{n-1}] + d_{n-2} - 2a_{n-1}}{2} \right]^+ \quad (12)$$

This result is consistent with equation (7) as $\frac{dh_{n-1}}{dh_{n-1}} = 1$ and $\frac{h_n^*}{dh_{n-1}} = 0$.

Suppose Equality (11) is true for every $q > i$, then since d_{q-i} is the only term impacted by h_{q-1} in Equation (11), for any $q^o > i$ we have:

$$\frac{dh_{q^o}^*}{dh_i} = \frac{1}{1 + \sum_{j=i+1}^n \frac{E[I_j^{q^o}]}{(j^o - q^o)}} \frac{dh_{q^o-1}^*}{dh_i} \quad (13)$$

Assuming we know that $I_{j^o}^i = 1$ for some $j^o > q$, we obtain:

$$\left[-\frac{dh_{q^o-1}^*}{dh_i} + \frac{dh_{q^o}^*}{dh_i} \right] = \left[-1 + \frac{1}{1 + \frac{1}{j^o - q}} \right] \frac{dh_{q^o-1}^*}{dh_i} = \frac{-1}{j^o - q^o + 1} \frac{dh_{q^o-1}^*}{dh_i} \quad (14)$$

By recursion we get:

$$\frac{dh_{q^o-1}^*}{dh_i} = \prod_{s=1}^{q^o-i} \frac{1}{1 + \frac{1}{j^o - q^o + s}} = \frac{j^o - q^o + 1}{j^o - i + 1} \quad (15)$$

Combining Equation (14) and (15), we obtain:

$$\left[-\frac{dh_{q^o-1}^*}{dh_i} + \frac{dh_{q^o}^*}{dh_i} \right] = \frac{-1}{j^o - i + 1} \quad (16)$$

By Corollary 3.3, it is known that:

$$h_i^* = \left[-a_i + d_{i-1} - \sum_{j=i+1}^n E \left[\frac{d[-h_{j-1}^* + h_j^*]}{dh_i} [A_j - A_{j-1} + h_j^* - h_{j-1}^*] | A_2, \dots, A_i \right] \right]^+ \quad (17)$$

Taking the conditional expectation with respect to I_q^i , we obtain:

$$h_i^* = \left[-a_i + d_{i-1} - \sum_{q=i+1}^n E[I_q^i] \sum_{j=i+1}^q E \left[\frac{-1}{q - i + 1} [A_j - A_{j-1} + h_j^* - h_{j-1}^*] | A_2, \dots, A_i, I_q^i \right] \right]^+ \quad (18)$$

We then bring all the h_i^* terms to the lhs and isolate by dividing both sides by its coefficient: ³

³Notice that A_j 's and h_j 's cancel out in the telescopic sum. Also, $h_n = 0$ by definition.

$$h_i^* = \left[\frac{\sum_{q=i+1}^n E[I_q^i] \frac{E[A_q - A_i | A_2, \dots, A_i, I_q^i]}{q-i} + d_{i-1} - a_i}{1 + \sum_{q=i+1}^n \frac{E[I_q^i]}{(q-i)}} \right]^+ \quad (19)$$

We have shown that Equation (11) is true for $i = n - 1$ and that if it is true for $i + 1$, then it is also true for i . Therefore, the induction hypothesis is validated. \square

The optimal bus dispatching policy is a fixed point because we can combine the $\frac{n^2-n}{2}$ equations from (10) and the $n - 1$ equations from (11) to solve the $\frac{n^2-n}{2}$ indicator probabilities and the $n - 1$ holds. Finding the fixed point is hard because there is no way to obtain $E[I_q^i]$ directly. We propose a streamlined dispatching policy in analytical form and provide a bound on its deviation from the optimal policy.

2.4.2 Proposed control method

In this section, we present the proposed bus dispatching policy that stems from the structural properties of an optimal policy derived in Subsection 3.1. We have expressed the bus dispatching policy that minimizes headway variance without buffer time in terms of the $E[A_i]$'s and of $E[I_j^i]$'s. In the proposed control method, we approximate the $E[I_j^i]$ terms by mimicking the decision process that would take place with *a priori* knowledge of bus arrival times. We begin by showing an optimal dispatching policy with perfect information about arrival times. We then use this policy to derive the proposed method, which approximates the optimal bus dispatching policy knowing the joint probability distribution of arrival times but not their realization.

First consider the bus dispatching problem with *a priori* knowledge of bus arrival times. At any point in time, the number of buses dispatched must be less than or equal to the number of buses that have arrived at the control point. Therefore the frequency of buses sent must always be less than or equal to the frequency of buses that have arrived. A policy that minimizes the sum of squared headways without using buffer time must consist of dispatching buses at the same rate as the lowest

predicted frequency of arriving buses. This way, even though the latest bus will arrive with a *big gap* (low arrival rate), it will be dispatched at the same headway as the preceding vehicles. Figure 2 illustrates this policy by means of cumulative number of arriving and departing buses from a control point as a function of time. The slopes of the arrival and departure curves are the rates of dispatch. Under the optimal policy, each vehicle preceding the *latest bus* is held and they are all sent at a constant rate.

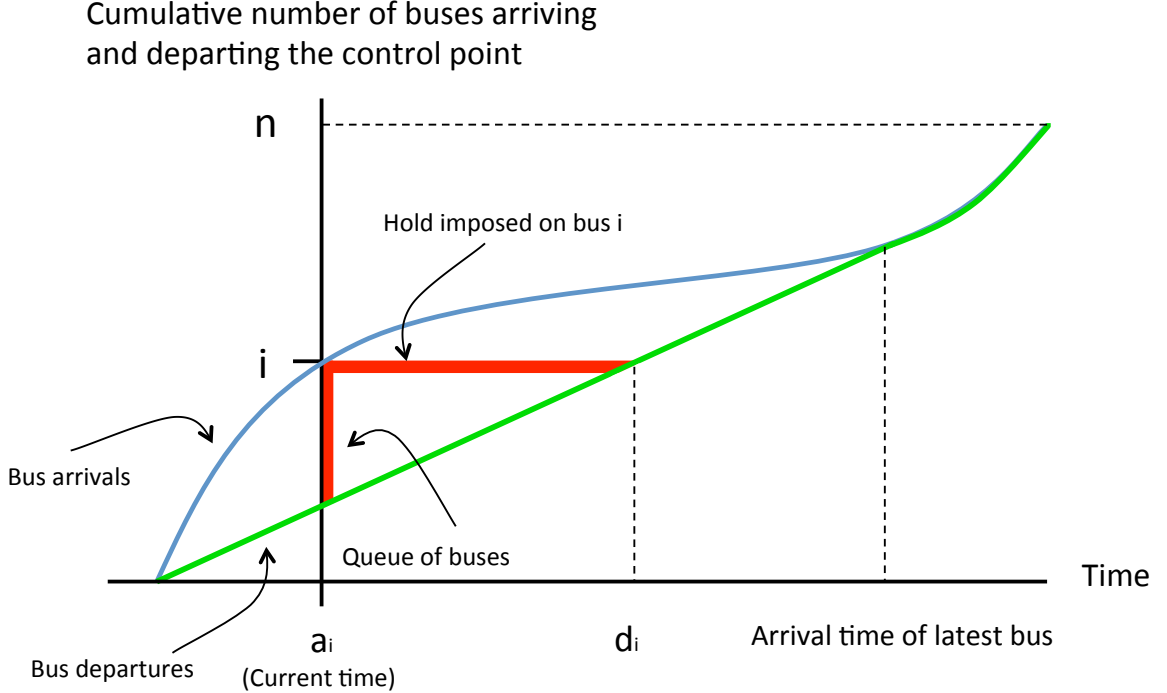


Figure 2: Bus arrival and optimal departure process at a control point

As shown in Proposition 2.4.6., an approximation to $E[I_j^i]$ (the probability that bus j is the latest after i) is the sum of its conditional expectations for every combination of arrival times. The approximation is inexact because it assumes that every hold from i to j will be made with complete information on the arrival times. The rhs of Equation (20) is the sum of conditional probabilities that bus j would be the first negative hold if every holding time had to be optimally determined at the i^{th} decision epoch, knowing the realizations of A_{i+1}, \dots, A_n , as per Figure 2.

Proposition 2.4.5.

$$E[I_j^i] \approx \sum_{a_{i+1}, \dots, a_j} E[I_j^i | A_{i+1} = a_{i+1}, \dots, A_n = a_n] P[A_{i+1} = a_{i+1}, \dots, A_n = a_n] \quad (20)$$

$$= P \left[j = \arg \max_r \frac{A_r - a_i}{r - i} \right] \quad (21)$$

The proposed method consists in identifying probabilistically the bus that will arrive the latest, and to hold each preceding bus to prevent the lagging bus from departing with a *big gap*. The hold imposed on bus i under the proposed dispatching policy is shown in Equation (22).

$$h_i^* \approx \left[\frac{E \left[\max_r \frac{A_r - a_i}{r - i} - (a_i - d_{i-1}) \right]}{1 + E \left[\left(\arg \max_{r-i} \frac{A_r - a_i}{r - i} \right)^{-1} \right]} \right]^+ \quad (22)$$

We will express the error of the approximation, ϵ , in terms of the variable ϵ'_j , which denotes the difference in conditional expectations for the inter-arrival times as follows:

$$\epsilon'_j = E \left[\frac{A_r - a_i}{r - i} | I_j^i \right] - E \left[\frac{A_r - a_i}{r - i} \right] \quad (23)$$

Corollary 2.4.6 justifies the choice of ϵ'_j 's as error terms. Its proof can be found in Appendix A.

Corollary 2.4.6. *If $\epsilon'_j = 0 \forall i$ then Equations (20) and (22) are equivalent to (10) and (11)*

Finally, we define $\epsilon'' = \max_j \epsilon'_j$, and because $\sum_{j=i+1}^n [I_j^i] = 1$, we get the following bound on the total error ϵ :

$$\epsilon \leq \frac{\sum_j E[I_j^i] \epsilon'_j}{1 + \frac{\sum_j E[I_j^i]}{j-i}} \leq \epsilon'' \quad (24)$$

Holds recommended by the proposed dispatching policy differ from the optimal by lesser order terms. The hold recommended by the proposed policy can be obtained by simply computing the expected arrival time and index of the bus running the latest. To test the performance of our proposed dispatching policy, we compare it with the optimal in a simple numerical example.

2.4.3 Numerical example

To evaluate the quality of our proposed method with respect to an optimal policy, we compare their performance in a simple numerical example. In this example, three buses are continually running on a loop-shaped route. Bus 1 just arrived at the control point, and the agent must decide how long it should be held. Bus 0 (also known as bus 3) left the control point five minutes ago. Bus 2 will arrive at time A_2 , and bus 3 will arrive at time A_3 . The arrival times of buses 2 and 3 are bi-variate normal random variables with a correlation parameter ρ . The probability distribution of A_2 and A_3 are shown below:

$$\begin{bmatrix} A_2 \\ A_3 \end{bmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \quad \boldsymbol{\mu} = \begin{bmatrix} 15 \\ 25 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 4 & 4\rho \\ 4\rho & 4 \end{bmatrix}$$

A dynamic programming approach was developed to find an optimal holding policy for the numerical example. The DP method evaluated the expected sum of squared headways for every combination of correlation parameter, ρ , bus arrival times, A_2 and A_3 , and decisions to hold bus 1 and bus 2 for non-negative amounts of time, h_1 and h_2 , assuming that bus 3 would not be held. The algorithm then selected the holding policy that minimized the expected sum of squared headways for each value of the correlation parameter.

The proposed method and the optimal policy recommended the same hold for bus 2 for every combination of correlation parameter, arrival times, and hold imposed on bus 1. However, the two methods recommended different holds for bus 1. Figure 3

shows three graphs with the correlation parameter, the hold imposed on bus 1 and the average passenger waiting time. The top left graphic shows the hold imposed by both policies for a wide range of correlation parameters. It can be seen that both methods yield almost the same passenger waiting time in the overlapping curves of the lower left graphic. The bottom right graphic shows average passenger wait in terms of h_1 , assuming h_2 is optimal, for three values of ρ . Although these curves are convex, a deviation of the top left graphic's y-axis range makes a very small difference in their average waiting time. As a result, the proposed policy is a very close approximation to the optimal policy.

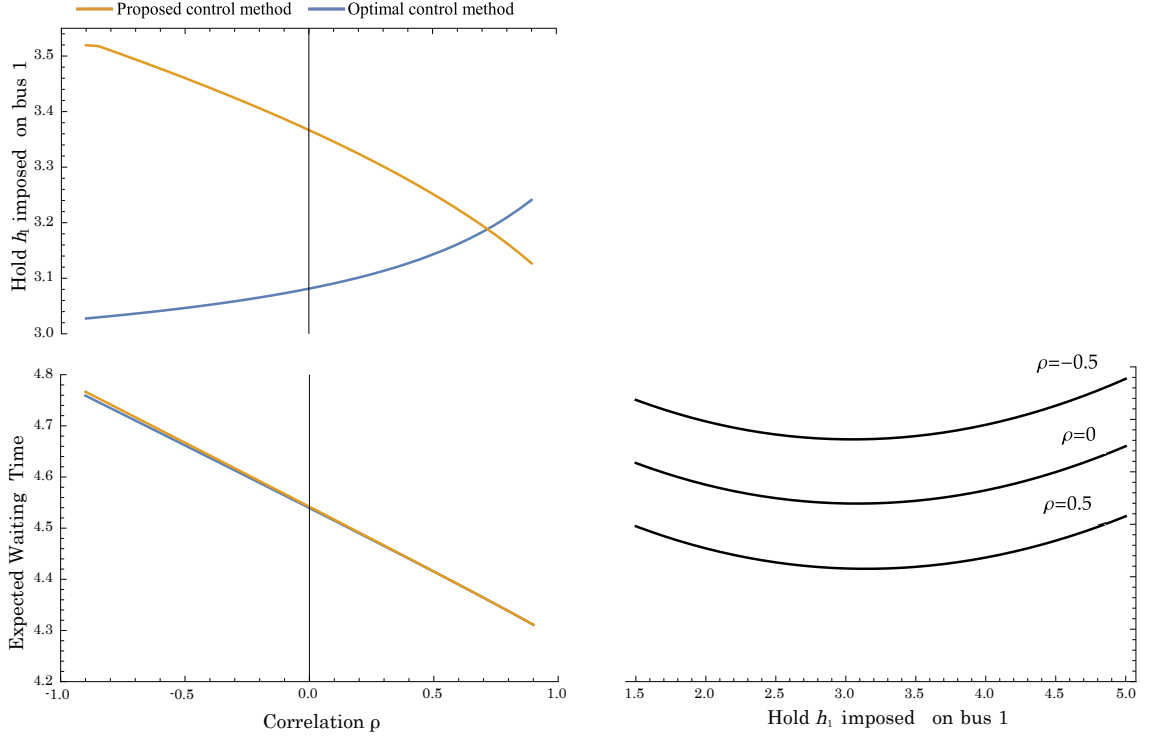


Figure 3: Optimal hold h_1 as a function of correlation between A_2 and A_3 .

2.5 Simulation

In this section, the proposed dispatching policy is compared through simulation with methods used in practice and recommended in the literature. In this example, seven buses run on a theoretical route with 25 stations, then return to the control point,

take a one minute break, and are re-dispatched according to the policy evaluated. Link travel time from one station to the next is normally distributed with mean of one minute and standard deviation of six seconds. At each station, there is a random stream of arriving passengers that have equal probability of alighting at each downstream station through a Poisson Process. The three dispatching policies tested are the Headway-Based policy, the Self-Coordinating policy, and the proposed control strategy. The purpose of the simulation is to compare the quality of service and resilience of each dispatching policy on a route with unstable headway dynamics.

The Headway-Based bus dispatching policy is a control method that aims to dispatch buses according to a *threshold* and without schedule (Daganzo, 2009; Abkowitz and Lepofsky, 1990; Van Oort et al., 2010). Buses are held until their headways reach the *threshold*, denoted β . In the simulations, we set β as the 50th and 85th percentile of cycle time over n . This predefined buffer avoids disrupting operations as recommended by the TCRP Report 113 (Furth et al., 2006). Headway-Based policies are used to dispatch buses on many Bus Rapid Transit and other frequent bus routes around the world. They are more resilient than Schedule-Based because a threshold headway is maintained, and no headway at departure is ever shorter than the threshold. The Self-Coordinating policy was proposed by Bartholdi and Eisenstein and is described in further details in the literature review (Bartholdi and Eisenstein, 2011). Each time a bus arrives at the control point it is either held for half of the expected time until next arrival, or until its headway reaches the *planned headway* β as in the Headway-Based method. The proposed control method was simulated as per Equation (22).

The Self-Coordinating method and the proposed method require predictions on future arrivals. We developed an embedded simulation to generate cases of future arrivals, and collected their histograms, which we treated as probability distributions. These distributions were then used to generate the expected arrival times of the next

bus for the Self-Coordinating method and of the next $n - 1$ buses for the proposed method.

The mean passenger waiting time at the first station was used as the main policy evaluation criterion. Since headway variance tends to grow along a route, mean passenger waiting time at the first station is characteristic of the overall system performance (Hickman, 2001). This criterion rewards both short and stable headways. The performances of the dispatching policies were evaluated in two simulation experiments. In the first one, the control variable is a dimensionless parameter for systemic instability. In the second experiment, a perturbation is activated after some time.

2.5.1 Instability

Boarding and alighting operations are the central source of headway instability (Milkovits, 2008). In this experiment, we define a dimensionless instability parameter equal to passenger rate of arrival, multiplied by boarding and alighting time. We found that for a fixed value of the instability parameter, any combination of its components yields the same passenger waiting time, mean headway and coefficient of variation. Figures 4, 5 and 6 show 95th percentile confidence intervals for the mean passenger waiting time, mean headway, and headway coefficient of variation respectively, after 20 vehicle dispatches have occurred.

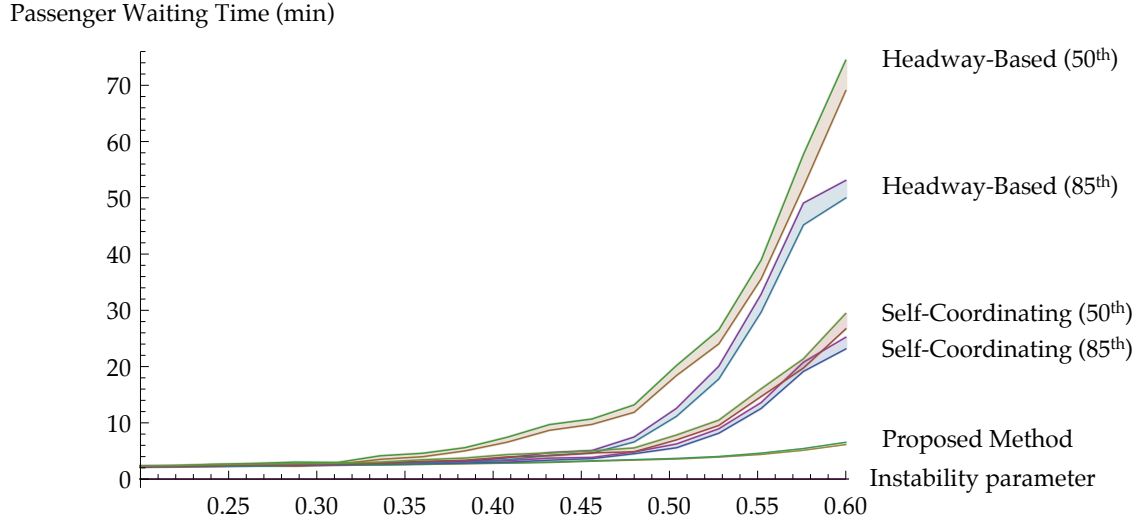


Figure 4: Mean passenger waiting time versus level of systemic instability

For low levels of systemic instability, the performance of each dispatching method is roughly equal. The headway coefficient of variation rises quickly in the Headway-Based method, causing passenger waiting time to increase at a high rate. The Self-Coordinating method keeps operations stable until the instability parameter reaches 0.5. The proposed control method maintains the coefficient of variation under 0.2. It is able to adapt to instability by increasing all headways and avoiding the formation of big gaps.

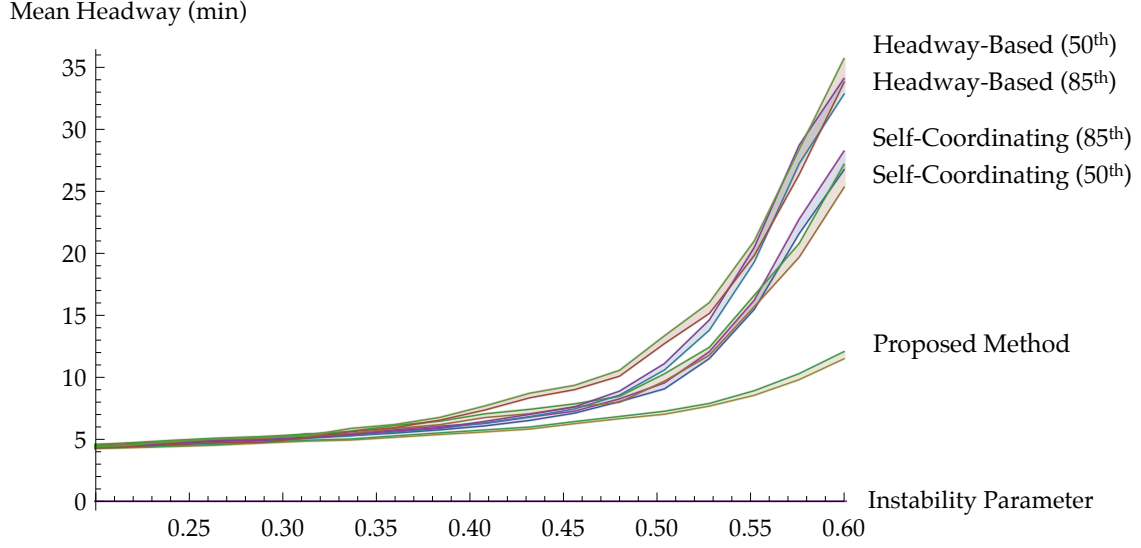


Figure 5: Mean headway versus level of systemic instability

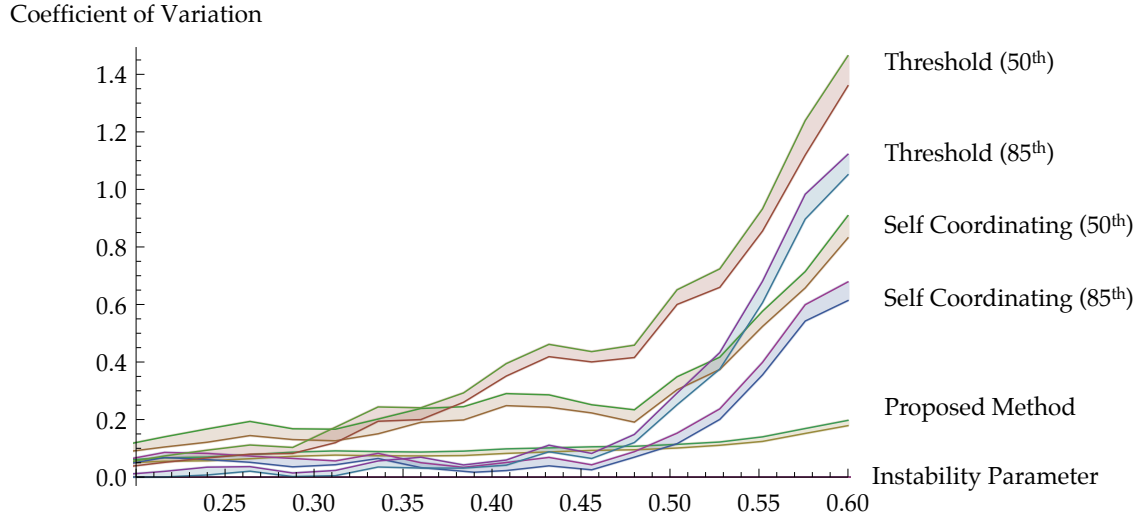


Figure 6: Headway coefficient of variation versus level of systemic instability

2.5.2 Perturbation

The Instability experiments were run in different operating conditions, but always in stationary environments. To test the resiliency of the dispatching methods, we have induced a perturbation in the following simulation. The route starts with a rate of 0.36 passengers arriving per station per minute for each origin-destination stream and

one second per boarding and alighting (dimensionless instability parameter of 0.36). When the 70th bus is dispatched, the rate of arriving passengers doubles until that bus gets back to the control point, then the system returns to its initial level of instability. Figure 7 shows a 95th percentile confidence interval for the mean passenger waiting time as vehicles are dispatched. The Headway-Based and Self-Coordinating methods were simulated using the 85th percentile of running time.

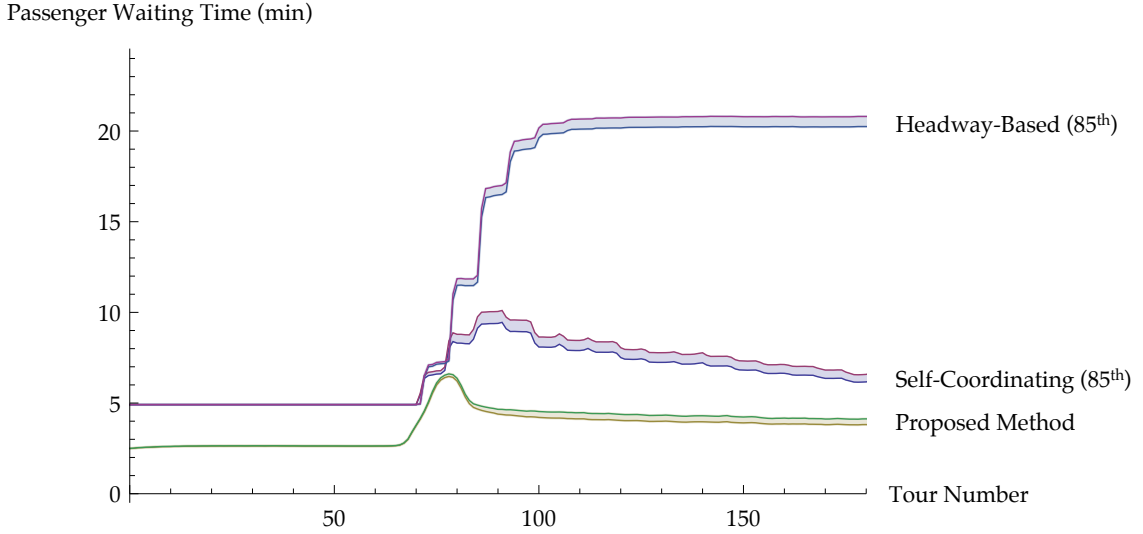


Figure 7: Mean passenger waiting time versus number of dispatches

Before the perturbation, the Headway-Based and the Self-Coordinating methods systematically dispatch vehicles at the threshold headway. The mean passenger waiting time under the proposed control method is shorter because headways are kept stable without planned buffer time. When the perturbation is introduced, the Headway-Based method undergoes severe disruptions, and passenger waiting time increases permanently. The Self-Coordinating and the proposed method forecast the perturbation before it occurs, and are able to adjust their operations accordingly. In the Self-Coordinating simulation, passenger waiting time increases initially, then starts to decrease as the dispatching method gradually controls *big gaps*. The proposed control method stabilizes the system within seven dispatches of the perturbation offset. Passenger waiting time is higher after the perturbation because more

passengers are introduced into the system and operations are slowed down. The method, however, maintains low headway variance at the highest frequency and the lowest mean passenger waiting time.

2.6 Conclusion

In this work, we presented a method to hold buses at control points using real-time information. The objective was to create a bus dispatching policy that would minimize the sum of squared headways at departure from the control points, assuming that the last bus to arrive would be dispatched immediately. In other words, we sought to minimize the waiting time of passengers arriving at the control point as a stationary Poisson process, while maximizing bus frequency. Maintaining the highest possible frequency is equivalent to maximizing the rate of bus dispatching; it makes controlling headway regularity difficult because in our problem, buses can be held but not accelerated. Therefore, a bus that arrives at a control point much later than the last departure will have to be dispatched with a *big gap* causing undue passenger wait. Our problem could be extended to use buffer time by assuming that the last bus could be held, but have the recourse to start a new route immediately upon arrival.

Our formulation of the bus dispatching problem takes an agency perspective and considers that if buses are dispatched at high frequency with stable headways, then the flux of arriving passengers will be evenly spread among the passing vehicles. Due to unstable headway dynamics, the variance of headway and passenger load increase monotonically along a route (Hickman, 2001). We did not explicitly consider the waiting time of passengers waiting at stations downstream of the control point because unstable headway dynamics are a global phenomenon; only by dispatching buses at stable headways can we prevent headway variance from increasing uncontrollably along the route.

The problem of dispatching buses on a high frequency route was addressed as a

stochastic decision problem. Assuming knowledge of the joint probability distribution of bus arrival times, we derived an optimal dispatching policy in terms of expected arrival times, and of the probability of holding following buses under an optimal policy. We proposed a dispatching method that approximates the optimal policy but that only needs expected bus arrival times and expected minimal rate of bus arrival. We found a bound on the approximation analytically. We also compared the proposed method with the optimal in a numerical example, and found that both methods produced very similar passenger waiting time. These results indicate that using bus arrival times and rate of arrivals to control a bus route, as per the proposed method, can produce almost the same results as an optimal dispatching policy that uses a joint probability distribution of bus arrival times.

We simulated a bus route with 25 stations and seven buses, to compare the proposed method with methods used in practice and recommended in the literature. We found that the proposed method yields shorter passenger waiting time under a wide range of operating conditions, and that it is capable of recovering from a perturbation much better than other methods.

Future research could test the proposed method on a real bus route and evaluate its effect on operations. Using real-time arrival predictions, the holds recommended by the proposed method could be computed automatically and communicated to bus operators through tablets, in the vehicles or at the control points. The tablets would tell operators how long to hold upon arrival at the control point and send a flashing signal at the time of recommended departure. Headways could be monitored via Automatic Vehicle Location; passenger waiting time could be measured by direct observation; and the overall satisfaction of passengers and operators could be surveyed.

CHAPTER III

COMPARING BUS HOLDING METHODS WITH AND WITHOUT REAL-TIME PREDICTIONS

3.1 *Abstract*

On high-frequency routes, transit agencies hold buses at control points and seek to dispatch them with even headways to avoid bus bunching. This paper compares holding methods used in practice and recommended in the literature using simulated and historical data from Tri-Met route 72 in Portland, Oregon. We evaluated the performance of each holding method in terms of headway instability and mean holding time. We tested the sensitivity of holding methods to their parameterization and to the number of control points. We found that Schedule-Based methods require little holding time but are unable to stabilize headways even when applied at a high control point density. The Headway-Based methods are able to successfully control headways but they require long holding times. Prediction-Based methods achieve the best compromise between headway regularity and holding time on a wide range of desired trade-offs. Finally, we found the prediction-based methods to be sensitive to prediction accuracy, but using an existing prediction method we were able to minimize this sensitivity. These results can be used to inform the decision of transit agencies to implement holding methods on routes similar to TriMet 72.

3.2 *Introduction*

On high frequency routes, there is a natural tendency for buses to *bunch* together. When a bus is traveling with a long headway¹, it has to pick up and drop off a

¹In this text, the term *headway* will be used as the time between the passing of two consecutive buses at a single location. Later, we will distinguish between the headway (or forward headway) and

relatively greater number of passengers, which slows down the bus even more. As the lagging bus becomes crowded, the following buses only have a few passengers to serve, which they can do relatively fast. Eventually the lead bus may get caught by one or several following buses and they start traveling as a platoon. Bus bunching is the product of unstable dynamics that cause delays to grow (Hickman, 2001). Even a small perturbation such as a traffic signal or a passenger paying in cash can destabilize the route and lead to bus bunching (Kittelson, 2003; Milkovits, 2008).

Unstable headway dynamics are a systemic problem that causes passenger wait and crowding. Fan and Machemehl (2009) showed that on routes where headways are less than 12 minutes, passengers tend to arrive randomly, even if a schedule is available. Because more passengers arrive during long headways than during short ones, gaps in service cause disutility to passengers in the form of undue waiting time and crowding (Newell and Potts, 1964; Milkovits, 2008). One way for transit agencies to stop the progression of instability among headways, is to provide control points, where buses with short headways can be held to absorb the delay of following buses.

Holding buses at control points can help reduce at-stop passenger waiting time, but it increases the wait of passengers who have already boarded. There is a trade-off between stabilizing headways and maintaining high operating speed (Furth et al., 2006; Furth and Wilson, 1981; Cats et al., 2011). This is why transit agencies value the benefit of headway reliability and the disadvantage of holding time differently.

In addition to selecting a holding method for their routes, transit agencies also need to decide how to implement it. Several holding methods in the literature require setting a parameter, which affects the trade-off between holding time and headway stability (Daganzo, 2009; Xuan et al., 2011; Bartholdi and Eisenstein, 2011; Daganzo and Pilachowski, 2011). Holding methods can also be applied at one or several control points along the route, which may impact the performance of each method.

the backward headway, which is the time until the following vehicle will reach the current location.

Understanding sensitivity of holding methods on the parameterization and number of control point is necessary to select the most adequate holding method based on route characteristics and desired trade-offs

Several methods are based on predictions for the arrival times of following buses (Bartholdi and Eisenstein, 2011; Daganzo and Pilachowski, 2011; Berrebi et al., 2015). The quality of the predictions may affect transit operators’ ability to leverage headway stability from holding time. The required level of prediction accuracy and confidence can be burdensome for certain transit operators that may not need high-quality predictions for other applications. The ubiquity of prediction-based methods is therefore dependent on their sensitivity to prediction quality.

Research in the literature has compared holding methods, but there currently lacks a unified framework to evaluate the conflicting objectives of stabilizing headways and minimizing holding time. Xuan et al. (2011) and Berrebi et al. (2015) have case study sections to compare methods in the literature to their own. Cats et al. (2011) compare naive methods used in practice and a headway-based method similar to the method in Daganzo and Pilachowski (2011). There is a need for a sensitivity analysis to support the choice of holding methods and their parameterization based on route characteristics, including the number of control points on routes similar to Tri-Met 72.

In this paper, we investigate the holding trade-off of holding methods used in practice and recommended in the literature. To this end, we evaluate holding methods on a simulated bus route using historical data from Tri-Met Route 72 in Portland, Oregon. We use the prediction tool developed in Hans et al. (2015) to reproduce the predictions in a realistic setting. In the following section, we describe the holding methods used in practice and recommended in the literature. In Section 3.4, we discuss the simulation experiment, and particularly the methods evaluated. In Section

3.5 we compare the performance of each holding method. In Section 3.6, we investigate the impact of parameter choice, and number of control points on the trade-off between stabilizing headways and keeping short holding times. In Section 3.7 we test the sensitivity of prediction-based holding methods on the accuracy and confidence of predictions. Finally, we provide concluding remarks in Section 5.

3.3 Holding methods in the literature

Methods to hold buses at control points have been addressed for many decades. Osuana and Newell (1972) and Newell (1974) formulated the theoretical basis for holding mechanisms to minimize passenger waiting time on simple routes in the 1970's. Since then, two main approaches to the bus holding problem have been developed in the literature, mathematical optimization and analytical.

The first approach consists in optimizing a weighted function of passenger wait in mathematical programs that consider the dynamics of bus trajectories (analytically or by simulation). At each decision stage, the optimization tools model the future states of the system, and assign holds on a rolling horizon. Hickman (2001) developed a linear search optimization algorithm which considers holding decisions in isolation of each other based on a stochastic model for bus trajectories. In Eberlein et al. (2001) a heuristic algorithm is used to minimize the waiting time of passengers at stops in a quadratic program. Bukkapatnam and Dessouky (2003) developed an iterative model where buses and stations *negotiate* holding time to minimize marginal costs. The method in Zolfaghari et al. (2004) assigns all holding decisions simultaneously, while considering capacity constraints, using AVL data and perfect predictions. Delgado et al. (2009) and Delgado et al. (2012) developed a simulation-based optimization algorithm that reproduces stationary bus trajectories deterministically and minimizes a weighted function of wait time. Sánchez-Martínez et al. (2016, 2015) extended their methods to consider dynamic passenger arrival rates and travel time. Cortés et al.

(2010) used a genetic algorithm to solve a multi-objective dynamic problem.

The second approach assigns holds as closed-form functions of bus arrival times (Daganzo, 2009; Daganzo and Pilachowski, 2011; Xuan et al., 2011; Bartholdi and Eisenstein, 2011; Berrebi et al., 2015). Buses are held with the objective of maintaining stable headways, and preventing bus bunching from the onset, which can minimize passenger waiting time globally and durably. Methods assign holds to buses as a function of the schedule, headways and, for some, predicted arrival times². Unlike the mathematical programming approach, these methods do not consider the prediction model further than its output. Therefore, they can use any prediction model³, which makes them much easier to implement. Closed-form holding methods will be the focus of this paper.

The notation used in this paper is consistent with Berrebi et al. (2015), and is shown in Table 2. The arrival time of the i^{th} bus at the control point is A_i (random variable) for a future arrival, and a_i (realization of A_i) for a known arrival time. Once the i^{th} bus arrives, it holds for time h_i , and is dispatched at time $d_i = a_i + h_i$. If the route runs according to a schedule, \bar{d}_i is the scheduled departure time from the control point, and H_i is its scheduled headway ($H_i = \bar{d}_i - d_{i-1}^-$). The holding methods evaluated in this paper are described in their Eulerian version (as in Xuan et al. (2011)) in Table 3 with the information they require and their recommended holding times.

3.3.1 Naive methods

The simplest and most widely used methods to hold buses at control points are to plan scheduled departure times, \bar{d}_i , or scheduled headways, H_i , well in advance (Boyle

²In the remainder of this text, we refer to holding methods that consider schedules as “schedule-based” and methods that consider headways as their main input as “headway-based”. Schedule-based methods include the Naive Schedule and the method recommended in Xuan et al. (2011). Headway based methods include the Naive Headway and the method recommended in Daganzo (2009).

³A discussion follows in Section 3.7.

Table 2: Summary of variable definitions.

Variable	Definition
A_i	Arrival time of bus i (Random Variable)
a_i	Arrival time of bus i (Realization)
h_i	Hold imposed on bus i
d_i	Departure time of bus i (Realization)
\bar{d}_i	Scheduled departure time of bus i
H_i	Scheduled headway of bus i
n_i	Number of following buses when bus i reaches the control point
CV^2	Headway Coefficient of Variation squared
$T[A_j]$	Arrival time of a particle ⁴ of the j^{th} bus

et al., 2009; Abkowitz and Lepofsky, 1990; Van Oort et al., 2010). In this paper, we will consider the Naive Schedule and Naive Headway methods as holding buses until their departure time or headway reaches the planned threshold, as shown in Equations 25 and 26 of Table 3. The Naive Schedule method is easy to implement because it only requires information about the arrival time of the vehicle being controlled. The Naive Headway method requires the last departure time from the control point. This method never dispatches buses with short headways because it imposes a threshold headway, H_i . This feature allows the Naive Headway method to control *big gaps* in service that follow buses with small headways.

3.3.2 Partial holding methods

Daganzo (2009) developed a headway-based holding method that compensates for unstable headway dynamics. A dimensionless parameter β accounts for the linear delay of vehicles resulting from a unit headway increase; values usually range between 0.01 and 0.1. When a bus arrives at a control point, its headway is readjusted to the scheduled headway, H , by a factor of $\alpha + \beta$, where $\alpha \in]0, 1[$. Equation 27 of Table 3 shows the hold imposed on a bus that arrives at a control point⁵.

⁵The forward headway in Daganzo (2009), Daganzo and Pilachowski (2011), and Xuan et al. (2011) is expressed in terms of inter-arrival time, $a_i - a_{i-1}$ without considering the hold imposed on the leading bus. To make these methods more robust, we have replaced the inter-arrival time by the time since last departure for the forward headway.

Xuan and collaborators then generalized this class of control, and developed a method that only considers the forward headway and deviation from schedule as shown in Equation 28 of Table 3 (Xuan et al., 2011). The method readjusts the headways with respect to the scheduled headway, H , by a factor β and the off-schedule time, $a_i - \bar{d}_i$, by a factor $\alpha \in]0, 1[$. The authors showed that the holding mechanism was capable of maintaining a schedule in a stochastic environment.

The partial holding methods act like parametric versions of the Naive Schedule and Naive Headway, with the α term in place to reduce the holding time. Partial holding methods rely either on the scheduled departure, \bar{d}_i , or the scheduled headway, H_i , to stabilize operations as in Naive methods. Unlike the Naive methods, however, they only recommend holding for a portion of the thresholds to reduce the holding time⁶. Daganzo, Xuan, and collaborators showed that their methods can recover from bounded deviations from the schedule or scheduled headway in stationary operating conditions. In practice, running time can be highly stochastic and non-stationary, which can cause systematic deviations from the schedule or scheduled headway. The methods described thus far do not consider the predicted arrivals of following buses to adjust target headways.

3.3.3 Prediction-based holding methods

A novel closed-form approach to the bus holding problem is to dispatch buses according to the predicted arrival times of following buses at the control point. Real-time prediction methods are becoming increasingly accurate and available, which allows the replacement of planned operations by the natural headway in current operating

⁶To preserve the robustness of their method, and recover from perturbations, Daganzo (2009), Daganzo and Pilachowski (2011), and Xuan et al. (2011) recommend using slack time. Slack time leverages longer holding time to stabilize headways, as shown in Argote-Cabanero et al. (2015a). We have found in a simulation, however, that adding slack time to the Route 72 schedule does not substantially affect the trade-off between headway stability and holding time. This is why we decided to calibrate slack time to the historical schedule in this paper.

Table 3: Holding methods with their data requirements and recommended holding times.

Holding method	Data requirement	Recommended holding time	Eq
Naive Schedule	Schedule	$\bar{d}_i - a_i$	(25)
Naive Headway	Forward headway	$H - (a_i - d_{i-1})$	(26)
Daganzo (2009)	Forward headway	$(\alpha + \beta)(H - (a_i - d_{i-1}))$	(27)
Xuan et. Al (2011)	Forward headway	$\beta(H - (a_i - d_{i-1})) - \alpha(a_i - d_i)$	(28)
Bartholdi and Eisenstein (2011)	Predicted backward headway	$Max[H - (a_i - d_{i-1}), \alpha(E[A_{i+1}] - a_i)]$	(29)
Daganzo and Pila-chowski (2011)	Forward and predicted backward headway	$(\alpha + \beta)(H - (a_i - d_{i-1})) - \alpha(H - (E[A_{i+1}] - a_i))$	(30)
Berrebi et. Al (2015)	Joint probability distribution of next n bus arrival times	$\frac{E\left[\max_{r-i} \frac{A_r - a_i}{r-i} - (a_i - d_{i-1})\right]}{1 + E\left[\left(\arg \max_{r-i} \frac{A_r - a_i}{r-i}\right)^{-1}\right]}$	(31)

conditions. Today, the vast majority of public transportation agencies in the United-States are capable of tracking their vehicles in real-time (Grisby, 2013). Using real-time vehicle locations, increasingly sophisticated prediction algorithms have surfaced recently (Chien et al., 2002; Cathey and Dailey, 2003; Jeong and Rilett, 2004a, 2005; Chen et al., 2005; Shalaby and Farhan, 2004; Gurmu and Fan, 2014; Sun et al., 2007; Mazloumi et al., 2011). Most notably, Hans et al. (2014, 2015) developed a prediction method specifically for the purpose of real-time control. Their prediction algorithm can generate probability distributions of arrival times. Based on these predictions, holding methods can consider following buses to equalize headways, without having to rely on planned operations.

Using real-time predictions, Bartholdi and Eisenstein (2011) developed a holding strategy that can stabilize headways without the need for planned operations. The method consists in holding each vehicle for the predicted time until the next arrival by a factor $\alpha \in]0, 1[$ as in Equation 29 in Table 3. When several buses arrive in close succession, however, the method sends the middle few uncontrolled. To prevent this, Bartholdi and Eisenstein added a minimum forward headway⁷. Otherwise, the method can split the burden on a *big gap* between two buses: each vehicle leaves the

⁷To let the method act primarily on backward headways, we set the minimal forward headway as $H_i/2$.

control point with the weighted sum between its forward⁸ and backward headways. Unlike the methods cited thus far, their method involves a mechanism that acts locally to scale headways to the rate of arriving vehicles.

Daganzo and Pilachowski used predictions on the next arrival time to blend the forward headway with the backward headway, as shown in Equation 30 of Table 3 (Daganzo and Pilachowski, 2011). The method considers the forward headway in the same way as Daganzo (2009), but it also subtracts the deviation of the expected time until the next arrival from the scheduled headway, $H - (E[A_{i+1}] - a_i)$, by a factor $\alpha \in]0, \frac{1}{2}[$. The holding method can reduce the difference between the forward and backward headways by holding buses for a weighted sum of that difference.

More recently, Berrebi and collaborators developed a method that takes a global approach by considering every bus on the route. The method is a generalization of Daganzo and Pilachowski (2011), without H and β , that considers n buses. The vehicle that will arrive with the maximum relative delay, $\max_{r-i} \frac{A_r - a_i}{r-i}$, is probabilistically identified and each preceding vehicle is held to absorb a share of that delay, as in Equation 31 in Table 3. When the lagging bus arrives at the control point, it can be dispatched with approximately the same headway as the leading few. The method can diffuse big gaps organically without the need for schedules, scheduled headways, or any kind of explicit slack time.

3.4 *Case study*

3.4.1 The route

Holding methods were compared by simulation using data on real bus trajectories from Tri-Met bus route 72 shown in Figure 8. Buses run in mixed traffic on the perpendicular 82nd Avenue and Killingsworth Street in Portland, Oregon. The scheduled

⁸More exactly the backward headway of the following bus, a headway ago.

headway alternates between seven and eight minutes in the afternoon peak. Historically, Route 72 had a bus bunching problem during peak hours (Berkow et al., 2007). Buses on route 72, however, are equipped with a Computer Aided Dispatching (CAD) system that would allow replacing the current schedule with real-time control. To evaluate the potential benefits and disadvantages of applying each real-time holding method described in the previous section, we have tested their performance in a case study.



Figure 8: Map of TriMet Route 72 TriMet (2016)

We used data available on the Portland Oregon Regional Transportation Archive Listing (PORTAL)⁹ to build the simulation framework. The online open platform provides Automatic Vehicle Location (AVL), Automatic Passenger Counts (APC), traffic signal settings, and loop-detector data for September 15th to November 15th 2011 on a part of route 72. The data covers the entire portion on 82nd Avenue and

⁹<http://portal.its.pdx.edu/Portal/index.php/fhwa>

ten blocks of Killingsworth Street (until 72nd Avenue) towards Swan Island.

In the study, we used historical data leading up to the first control point ¹⁰. Each method had the opportunity to hold vehicles at that point to stabilize operations and mitigate bus bunching. Whenever the boarding and alighting times exceeded the hold recommended by a control method, buses were dispatched after they finish loading and unloading. The headways of buses leaving the control point were recorded to evaluate the performance of the dispatching strategy. We used simulated data downstream from the control point to take into account the impacts of each holding method on headway dynamics.

In this paper, we considered a single value of H_i : its historical value. In the historical data, vehicles arrive at the control point(s) according to a set scheduled frequency of service. The scheduled frequency at a control point determines the scheduled headway, because the frequency of departure cannot be greater than the frequency of arrival. If we reduced the scheduled headway, the holding methods would be unable to control buses at all. If we increased the scheduled headway, a continually accumulating queue of buses would form. This is why we did not test holding methods with varying values of H_i .

3.4.2 Prediction

Once buses arrive at the control point, each method can recommend a holding time based on the last departure time (recommended by the method) and the predicted next arrivals. The methods recommended in Daganzo and Pilachowski (2011) and Bartholdi and Eisenstein (2011) need the expected next arrival time, $E[A_{i+1}]$. The method recommended in Berrebi et al. (2015) needs the probability distribution of each future arrival to infer the maximum relative delay, $\max \frac{A_r - a_i}{r - i}$, and its corresponding index, $\arg \max \frac{A_r - a_i}{r - i}$.

¹⁰Several holding point are considered in Section 5.2.

The prediction tool used in our simulation is a particle filter combined with an event-based mesoscopic model developed in Hans et al. (2014) and Hans et al. (2015). The prediction tool is capable of generating simulated trajectories solely using vehicle location data. The predicted arrival times of buses at the control point are generated iteratively as a function of their dwell and travel times, as shown in Table 4. The dwell time is estimated as the sum of boarding, alighting and door operation time, assuming passengers board and alight through the same door. The number of passengers boarding are assumed to follow a Poisson distribution, with no capacity constraints. The share of passengers alighting follows a binomial distribution with passenger loads estimated using historical headways. Travel times between stations are generated as a Gamma distribution. To ensure that the particle filter accurately reproduces operations on Route 72, the parameters of dwell and travel times, such as the rate of arriving passengers and the mean travel time between stops were calibrated on historical data from the route. Odd days were used for the calibration, and even days were kept for the simulation.

Table 4: Particle generation

Equation	Variable	Definition
$A_i = \sum_{s=s_{\text{last}}}^{s_{\text{ctrl}}-1} \delta_{i,s} + \pi_{i,s}$	$s_{\text{ctrl}}-1$	Upstream stop from the control point
	s_{last}	Last stop visited by bus i
	$\delta_{i,s}$	Dwell time of bus i at stop s
	$\pi_{i,s}$	Travel time of bus i between stop s and $s + 1$
$\delta_{i,s} = aN_A + bN_B + c$	a	Individual alighting time
	b	Individual boarding time
	c	Time lost in door opening and closing
	N_A	Number of alighting passengers
	N_B	Number of boarding passengers
$N_B \sim P(d_s h_{i,s})$	$h_{i,s}$	Headway of bus i at stop s
	d_s	Demand rate at stop s
$N_A \sim \text{Bin}(L_{i,s}, \mu_s)$	$L_{i,s}$	Load of bus i at its departure from stop s
	μ_s	Alighting proportion at stop s
$L_{i,s} = (1 - \mu_s)L_{i,s-1} + d_s h_{i,s}$		
$L_{i,0} = 0$		
$\pi_{i,s} \sim \text{Gam}(\text{mean}_s, \text{stdev}_s)$	mean_s	Mean travel time between stop s and $s + 1$
	stdev_s	Standard deviation of travel time between s and $s + 1$

When a bus arrives at the control point, 100 particles (simulated bus trajectories) are generated for each following bus traveling on the route.¹¹ The arrival times of following buses at the control point are then aggregated in a histogram, which are treated as probability distributions. Figure 9 shows a time-space diagram with the

¹¹We chose to generate 100 particles as a compromise between the computational time and the resolution of the histogram.

trajectories of following vehicles in part (a), and the associated histogram of bus arrival times in part (b). At each station, the particles in part (a) tend to divert away from their mean because the prediction accounts for the delay accumulation caused by unstable headway dynamics. In part (b), the arrival time of the current and leading buses are represented by vertical lines because their arrival times are known. The arrival times of following vehicles are random and their distributions widen with the horizon of their prediction.

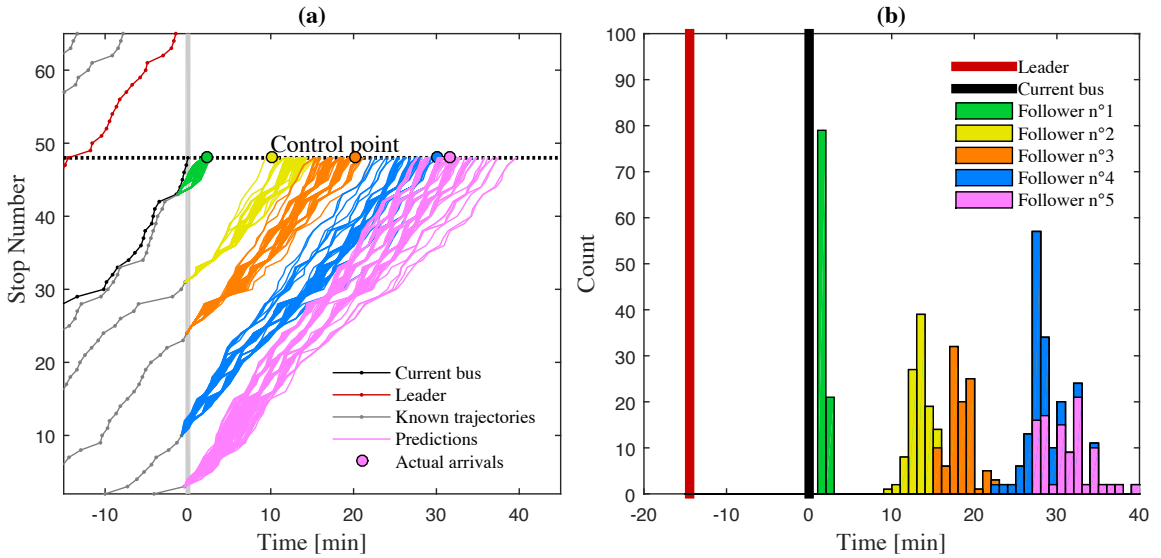


Figure 9: (a) Generation of particles to simulate bus trajectories. (b) Histograms of simulated bus arrival times treated as probability distributions.

The particle filter recommended in Hans et al. (2014) and Hans et al. (2015) is particularly suitable for real-time control. Unlike regression and machine learning prediction methods, the particle filter is capable of generating probability distributions, which is necessary for the method in Berrebi et al. (2015). The model is simple to calibrate, and it is compatible with any bus route and any data format. The tool can consider traffic congestion and traffic signal data if available, but in this study, we assumed that only vehicle locations were available for the prediction

3.4.3 Performance indicators

Performance indicators allow transit agencies to identify and address gaps in service. On high frequency routes, a measure of instability is the squared coefficient of variation of headways, denoted CV^2 as shown in Equation 32 (Kittelson, 2003). The variable measures the extent of headway instability. When $CV^2 = 0$, buses are evenly spaced, and when $CV^2 = 1$, Average Passenger Wait (APW) is equal to H . Assuming Poisson arrivals, APW is directly proportional to CV^2 (Newell and Potts, 1964), as shown in Equation 33.

$$CV^2[\text{headway}] = \frac{Var[\text{headway}]}{E[\text{headway}]^2} \quad (32)$$

$$APW = \frac{E[\text{headway}] (1 + CV^2[\text{headway}])}{2} \quad (33)$$

Variation in headways tends to increase along the route unless buses are controlled (Hickman, 2001). Therefore, unstable headway dynamics can have lasting effects on the system and on passenger experience if headways are not stabilized. We choose to evaluate headway stability in terms of CV^2 because it is a dimensionless parameter that directly relates to at-stop passenger wait and allows extrapolation of results to other routes.

Holding methods can stabilize headways by trading off holding time. Holding time too has a cost. It causes passengers who are already aboard the vehicle undue waiting time. In addition, holding time reduces bus operating speed. Finally, holding at a control point can disrupt surrounding traffic, depending on its location. Since transit agencies may value the benefit of headway stability and the cost of holding time differently depending on the route, we chose to keep these measures of performance separate.

3.5 *Cross-comparison*

We simulated the decision process of each holding method described in Table 3 in the afternoon peak hour, when scheduled headways, H_i , oscillate between seven and eight minutes. In the simulation, buses were held at the 48th station, which is seven miles away from the departure point, at the intersection with the light-rail, MAX, on Banfield Expressway. We chose this station because it is used as a schedule recovery point in historical data. The station also has by far the greatest alighting proportion and boarding demand on Route 72. These considerations are important because the overall cost of holding time is proportional to the number of passengers who ride through the holding point and the overall benefit of stabilizing headways is proportional to the number of passengers waiting at downstream stops. Therefore, holding at station 48 inconveniences few passengers and benefits many.

Every method was parameterized with β calculated at each arrival as per Daganzo (2009), Daganzo and Pilachowski (2011), and Xuan et al. (2011), and $\alpha = \frac{1}{2}$ to provide a middle-ground basis of comparison with the Naive Schedule and Headway methods. We discuss the choice of the α parameter in the next section.

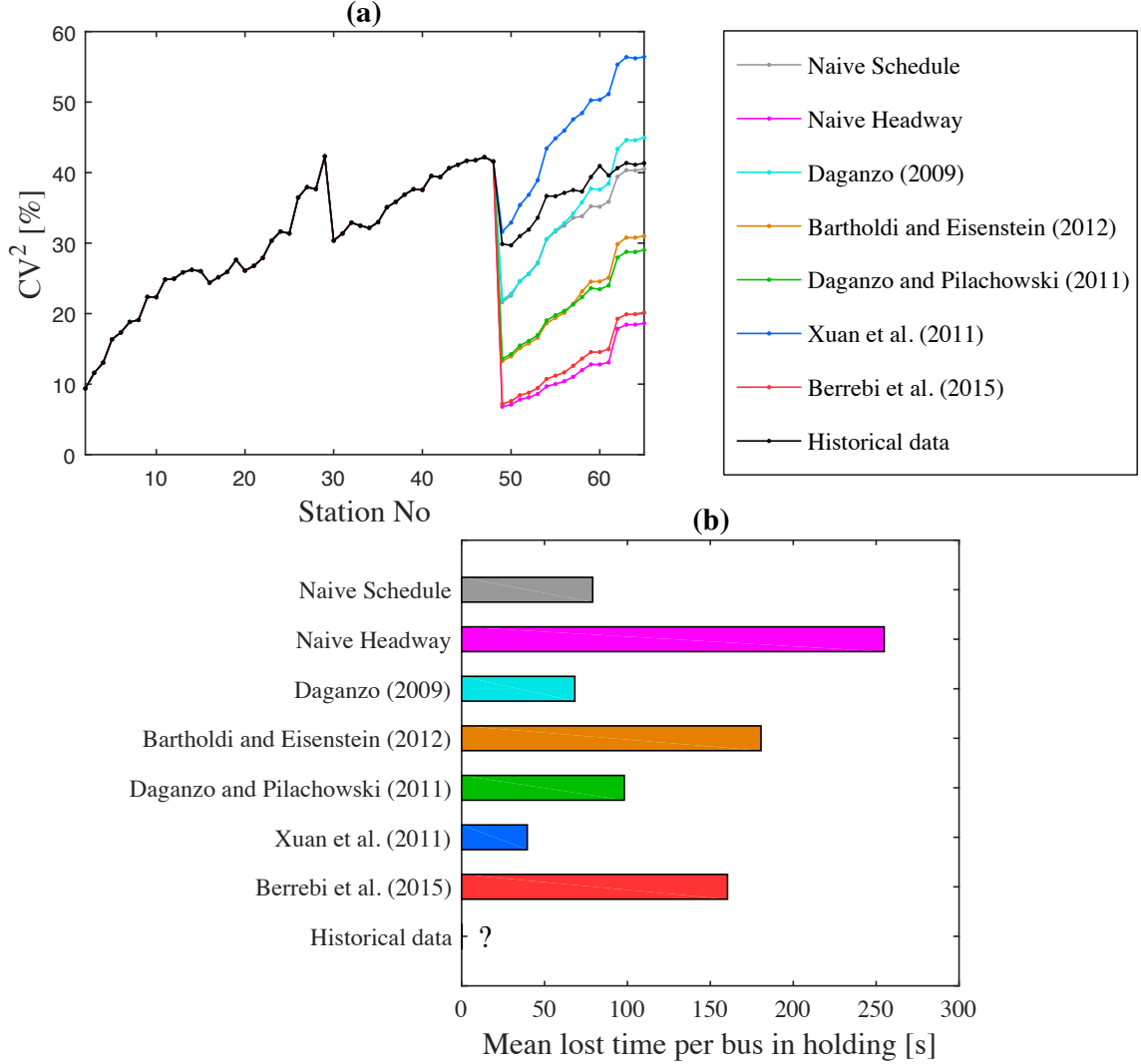


Figure 10: Performance of holding methods in terms of (a) squared headway coefficient of variation, (CV^2) and (b) mean holding time past boarding and alighting (lost time).

The results of our simulation are shown in Figure 10, with squared headway Coefficient of Variation, CV^2 in part (a) and mean holding time past boarding and alighting (lost time) in part (b) for each holding method. This figure and all figures following are based on 50 simulation runs. In part (a), values of CV^2 upstream of the control point come from historical data, and values downstream of the control point

are simulated, except for the *historical data* curves in black.

In historical data, buses start the route with close to even headways, but get destabilized along the route, leading to the control point. The CV^2 increases in a saw tooth pattern, with small drops corresponding to mid-route control points. When buses arrive at the main control point, CV^2 is 42%, which, according to the second edition of the Transit Capacity and Quality of Service, corresponds to a Level of Service (LOS) E, and is denoted as *Frequent bunching* (Kittelson, 2003). In historical data, CV^2 is reduced to 30%, which is still LOS E. We did not display the mean lost time in *historical data* because we could not differentiate the share that was conscious holding from boarding, alighting, and other dwell operations.

The holding method applied in historical data, and the method recommended in Xuan et al. (2011) are both based on the Naive Schedule method, but they produce different results. The Naive Schedule method¹² reduces CV^2 to 22% (LOS D, *Irregular headways, with some bunching*) with a mean lost time holding of 79 seconds. The holding method in Xuan et al. (2011) sends buses with greater variability than the Naive Schedule, scoring LOS E, but only holds buses for 40 seconds on average.

The Naive Headway method is capable of sending buses at regular intervals, but requires long holding times. The Naive Headway method was effective at stabilizing headways with 7% of CV^2 (LOS B, *vehicles slightly off headway*). The method, however, holds buses for 255 seconds on average. The holding times tend to accumulate because each late bus pushes back the dispatching time of all upstream buses. The method proposed in Daganzo (2009) dispatches buses with much more erratic headways than the Naive Headway, which causes longer wait for passengers at a stop, but the method keeps mean holding time at 40 seconds.

The prediction-based holding methods consider the predicted arrival times of following buses to diffuse big gaps. The method in Bartholdi and Eisenstein (2011)

¹²Sometimes overlapping with Daganzo (2009) in parts (a) and (b).

dispatched buses at $CV^2 = 0.13$ (LOS C, *vehicles often off headway*). The method, however, holds buses for 180 seconds on average. The method in Daganzo and Pilachowski (2011) produced almost the same results as in Bartholdi and Eisenstein (2011), but only required 98 seconds of holding time. The method recommended in Berrebi et al. (2015) considers the arrival time of every following bus on the route. The holding method reduces CV^2 to 7% (LOS B, *Vehicles slightly off headway*) with 160 seconds of mean holding time.

The reduction in headway variability at the control point has lasting effects downstream from the control point. The more a method is able to reduce headway variability at the control point, the less headways tend to destabilize further down the route. Conversely, the rate of increase of CV^2 is highest for methods that are the least able to stabilize headways. For example, between stops 49 and 60, CV^2 increased three times more for the method in Xuan et al. (2011) than the method in Berrebi et al. (2015), which dispatched buses with far more stable headways. This trend is not seen in historical data because it benefits from mid-route control points, which we have omitted in the simulation.

3.6 Sensitivity

The performance of control strategies can vary depending on how they are applied. It is important to understand the implications of parameter choice and the number of control points. In this section, we investigate how the performance of each method is affected by these factors.

3.6.1 Parameterization

The methods recommended in Daganzo (2009), Daganzo and Pilachowski (2011), Bartholdi and Eisenstein (2011), and Xuan et al. (2011) all require setting an α parameter, but their interpretation of the parameter is different. For the methods prescribed in Daganzo (2009) and Xuan et al. (2011), α corresponds to the fraction

of the deviation from headway or schedule that is to be recovered by the hold. When α is close to zero, buses are barely controlled, and for values α close to one, the holding methods are similar to the Naive Schedule and Headway. For the method recommended in Bartholdi and Eisenstein (2011), α is the fraction of the predicted backward headway that buses should hold for. The mean holding time is α times the average headway. In Daganzo and Pilachowski (2011) the α parameter weights the difference between the forward and backward headway. For values of α close to zero, more importance is given to the forward headway, and for values close to one, more importance is given to the backward headway.

Figure 11 shows (a) CV^2 as a function of α , (b) mean lost time as a function of α and (c) CV^2 as a function of mean lost time immediately downstream of the station 48 control point¹³. Solid lines represent the range of α parameter recommended by the authors of the holding methods, and dashed lines represent values of α outside the recommended range. Note that when α is greater than its recommended range, Daganzo (2009), Daganzo and Pilachowski (2011) and Xuan et al. (2011) compensate for perturbations excessively by dispatching buses past \bar{a}_i or H_i . In the last section, we set $\alpha = \frac{1}{2}$ for each parametric method. The solid dots on Figure 11 show the performance of each holding method as parameterized in the previous section.

¹³Note that the interpretation of α differs for each method.

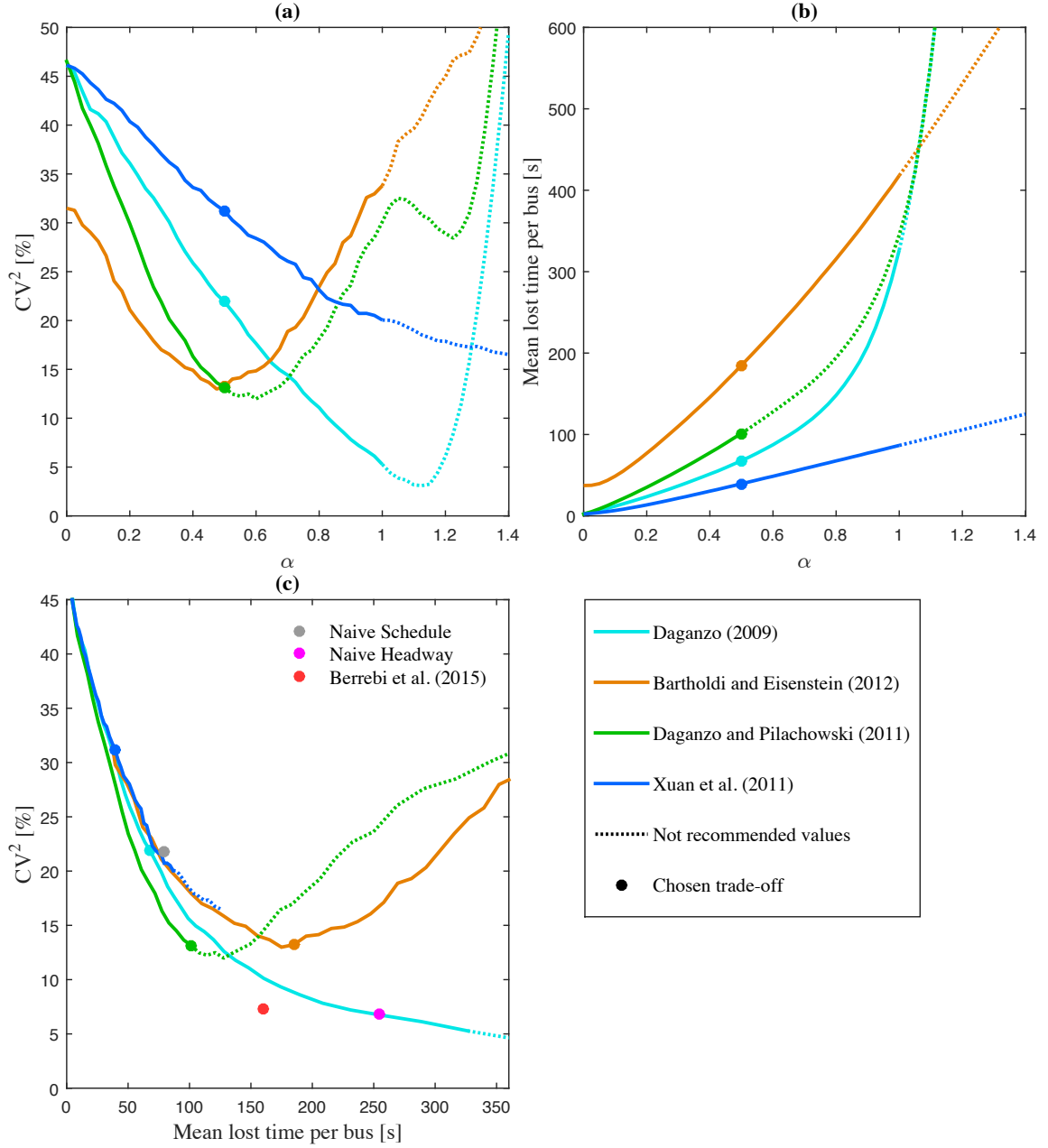


Figure 11: (a) CV^2 as a function of α , (b) mean lost time as a function of α and (c) CV^2 as a function of mean lost time at station 49.

The choice of α affects the amount of holding time recommended by each method, and ultimately the CV^2 and average passenger wait. Part (a) of Figure 11 shows that CV^2 for the partial holding methods decreases monotonically with values of α within their recommended range. Part (a) also shows that the parametric prediction-based

methods have convex shapes and attain their lowest CV^2 close to $\alpha = \frac{1}{2}$, with a slight deviation due to the β parameter. In part (b) of Figure 11, the holding time of each parametric method grows with α .

Part (c) of Figure 11 shows the trade-off between CV^2 and mean lost time attained by each method. The naive methods and the method recommended in Berrebi et al. (2015) are represented as single points because they are not parametric. The CV^2 in Xuan et al. (2011) and Daganzo (2009) decrease monotonically as a function of mean lost time. In both methods, the chosen trade-offs ($\alpha = 0.5$) yield far greater CV^2 than the naive methods from which they are derived. For the method in Xuan et al. (2011), the rate of decay remains high until the Naive Schedule trade-off is reached, whereas the method in Daganzo (2009) requires less than a third of Naive Headway's mean lost time.

Prediction-based methods achieve the best compromise between headway regularity and holding time in a wide range of settings. The method in Daganzo and Pilachowski (2011) can be parameterized to yield a lower CV^2 than any other method for any holding time up to 130 seconds. The method in Berrebi et al. (2015) can dispatch buses with 7% of CV^2 and 160 seconds of holding time, making it the preferable method for holding times over 130 seconds.

3.6.2 Multiple Control Points

Whereas methods thus far have been applied at a single control point, we now test them at several control points along the route. Applying holds at several control points can help maintain stable headways throughout, but it requires more frequent holding. In this section, we evaluate the trade-offs between headway stability and mean holding time for each method based on the number of control points. This analysis can support the decision of transit agencies to implement the holding method most adequate for their route and objectives.

The methods in Berrebi et al. (2015) and Bartholdi and Eisenstein (2011) are designed to dispatch buses with one or few control points. The method in Berrebi et al. (2015) considers the predicted arrival times of each bus on the route. It cannot be applied at a close succession of control points because holds would interfere with predictions. The method in Bartholdi and Eisenstein (2011) holds vehicles for αH_i on average. Holding time would therefore accumulate proportionally to the number of control points. Both methods, however, can be paired with on-route holding methods that only consider the local headway dynamics. We introduce two hybrid methods that apply the method in Bartholdi and Eisenstein (2011) (Hybrid #1) and Berrebi et al. (2015) (Hybrid #2) at a “main control point” (stop 29) and the method in Daganzo and Pilachowski (2011) at all other control points.

Figure 12 shows the mean CV^2 and total lost time applied at 1,2,4, and 8 control points for all methods and at 16 and 32 control points for the Naive Schedule and the methods in Xuan et al. (2011).¹⁴ ¹⁵ The α parameter was set to 0.5 for all methods, including both Hybrid Methods.¹⁶ The Hybrid Methods cannot be applied from the start of the route because they require the predicted arrivals of one or several upstream buses. All other methods, however, can be applied from the start of the route, where headways are more stable than anywhere else. We therefore decided to apply the Hybrid Methods in the second half of the route and all other methods throughout the route. In every case, control points were placed at regular intervals.¹⁷

¹⁴The two latter methods were the only ones applied at 16 and 32 control points because they are the only ones that do not consider headways. All other methods use either the forward or the backward headway, which would be affected by holds imposed at intermediate bus stops.

¹⁵Note that the β parameter is accumulated over the number of stops between control points for the method in Xuan et al. (2011). We did not, however, use a cumulative β when applied at a single control point.

¹⁶Testing was done with various values of α . We found that the partial methods require less holding time but are unable to stabilize the route with lower values of alpha.

¹⁷The control points selected were stops $\{30\}$, $\{30, 45\}$, $\{30, 38, 46, 54\}$, $\{30, 34, 38, 42, 46, 50, 54, 58\}$ for the Hybrid Methods and $\{30\}$, $\{20, 40\}$, $\{12, 24, 36, 48\}$, $\{7, 14, 21, 28, 35, 42, 49, 56\}$, $\{4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, 64\}$, $\{2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64\}$ for other methods.

Similarly, the CV^2 was averaged over the second half of the route for the Hybrid Methods and over the entire route for all other methods.

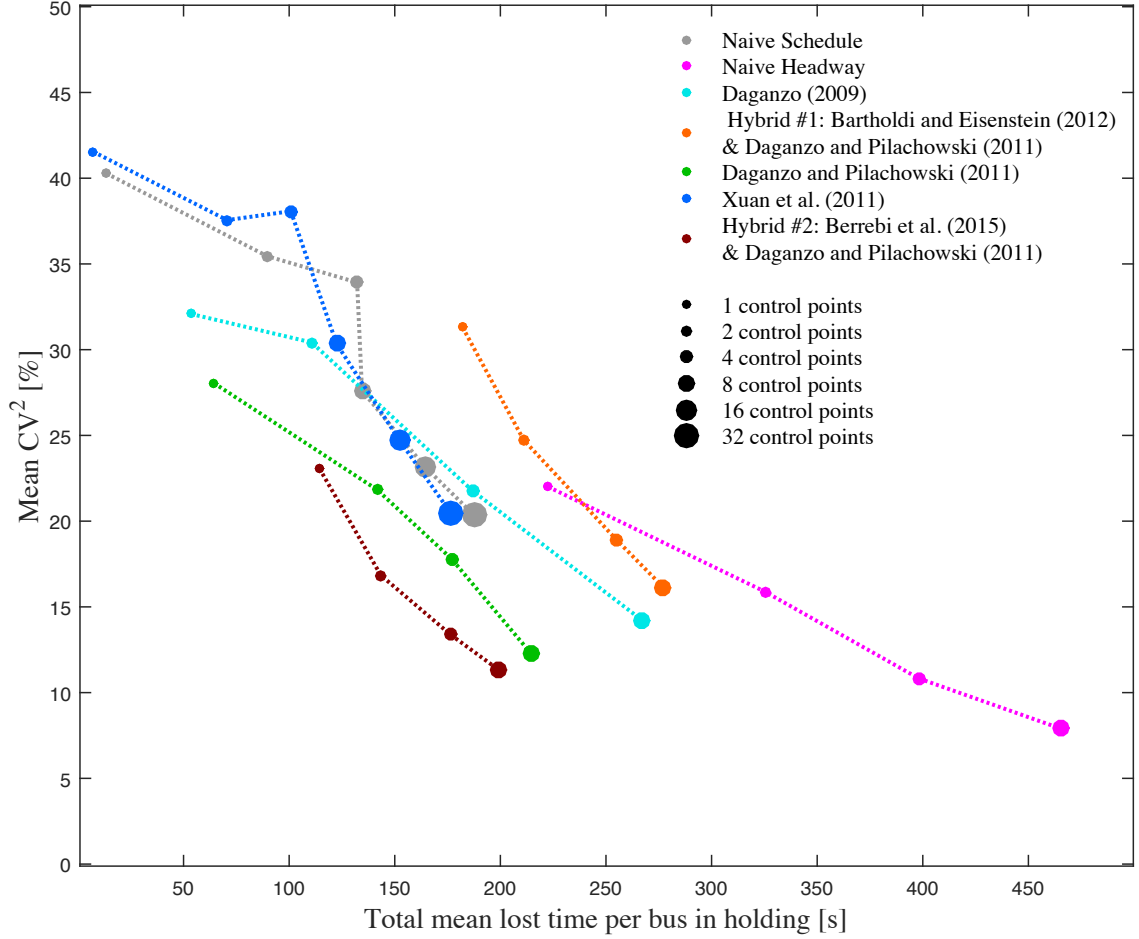


Figure 12: Mean CV^2 and total lost time applied at 1,2,4, and 8 control points for all methods and at 16 and 32 control points for the Naive Schedule and the methods in Xuan et al. (2011).

The number of control points affects the nature of the trade-off between holding time and headway stability. When applied at a single control point, the holding time is a riding cost imposed on passengers riding through. When applied at many control points, holding time can be approximated as a cost per distance. In any case, the at-stop waiting time due to uneven headways is a boarding cost for any passenger

getting on the vehicle.

The Schedule-Based Methods are similarly affected by the number of control points. The Naive Schedule and the method in Xuan et al. (2011) are unable to stabilize the route when applied sparsely. When applied at many stops (16 or 32) the methods require between 152 and 188 seconds to maintain a CV^2 between 20% and 25%, which is more than the methods in Daganzo and Pilachowski (2011) and Berrebi et al. (2015) for the same level of stability. As expected, the method in Xuan et al. (2011) yields more unstable headways but requires less holding time than the Naive Schedule method for any number of control points.

For any number of control points, the method applied in Daganzo (2009) requires much less holding time than the Naive Headway, but it also dispatches vehicles with greater CV^2 . The method recommended in Daganzo (2009) requires the same amount of holding time as the methods in Daganzo and Pilachowski (2011) and Berrebi et al. (2015) but yields greater CV^2 . When applied at 1,2, or 4 control points, the Naive Headway method yields the same level of stability as methods in Daganzo and Pilachowski (2011) and Berrebi et al. (2015) but requires far greater holding time. The Naive Headway can exceed the method in Berrebi et al. (2015) by 4%, when applied at 8 control point, at the cost of 266 additional seconds of holding time.

There is a discrepancy between Prediction-Based methods. The Hybrid # 1 yields both greater CV^2 and greater holding time than Partial Methods and other Prediction-Based methods for any number of control points. The method recommended in Daganzo and Pilachowski (2011) and the Hybrid # 2, on the other hand, yield the lowest CV^2 for any value of holding times between 114 and 398 seconds. The Hybrid # 2 dispatches vehicles with more stable headways than the method in Daganzo and Pilachowski (2011) for any number of control points and requires less holding time when applied at four or eight control points.

Figure 12 shows that for any value of holding time less than 177 seconds, the

combination of method and number of holding points yielding the lowest CV^2 , (a) has the greatest number of holding points the method can support under the set holding time, and (b) has the lowest number of holding point of any method under the set holding time. In other words, the best compromise between headway stability and holding time is always the method with the most concentrated holding time. In particular, the Hybrid #2 yields the best compromise between the two objectives but concentrates holding time in few holding points located in the second half of the route. The concentration of holding time disproportionately impacts the passengers riding through the few holding points. Transit agencies should therefore be mindful of finding holding points where many passengers board and alight or selecting methods that dilute holding time, even at the cost of more holding and instability.

3.7 Prediction

For transit agencies that wish to take advantage of bus dispatching methods using real-time predictions such as Daganzo and Pilachowski (2011), Bartholdi and Eisenstein (2011), and Berrebi et al. (2015), it is important to know what type of prediction is required and how accuracy will affect their performance. Although real-time vehicle tracking and prediction technologies are widely available among transit agencies, many of these systems were designed for passenger information rather than operational control. On these systems, using inadequate predictions could negatively impact the quality of dispatching mechanisms. Conversely, implementing a separate prediction system for control would duplicate efforts. To help transit agencies decide on the most appropriate prediction model, with respect to the performance of each holding method and to the costs of acquiring high-quality predictions, we tested the sensitivity of holding methods to the accuracy of predictions.

To evaluate the sensitivity of prediction-based holding methods to the prediction accuracy, we simulated their performance with synthetic predictions. Each time a bus

arrived at the control point, a synthetic distribution of arrival times of the following buses was generated with errors on the distribution mean, Δ_1 , and within the distribution, Δ_2 . The systematic error Δ_1 affects each prediction-based holding method because it biases the expected arrival time of each bus. The shape parameter on the other hand should not affect the methods in Daganzo and Pilachowski (2011) and Bartholdi and Eisenstein (2011) because they only require expected arrival times, but it may affect the method in Berrebi et al. (2015), which uses probability distributions.

Equations 34, 35, and 36 show the probability distribution of a synthetic trajectory of the j^{th} bus, denoted $T[A_j]$, at time a_i . The distribution is centered around $a_j + \Delta_1$, which is a uniformly distributed random variable, with interval length proportional to the horizon, $a_j - a_i$, and to the accuracy indicator, ϵ . The random variable Δ_2 is the error of trajectories around Δ_1 , which is normally distributed with scale parameter proportional to the horizon and to the confidence parameter, σ . The choice of Uniform and Normal distributions with the horizon, $a_j - a_i$ as parameters is consistent with the distribution of errors on the particle filter used in Sections 3.5 (Hans et al., 2015). Note, however, that the random variable Δ_1 is determined once for all synthetic simulated arrival times of the same bus, whereas Δ_2 is re-determined for each particle.

$$T[A_j] = a_j + \Delta_1 + \Delta_2 \quad (34)$$

$$\Delta_1 \sim U[-\epsilon(a_j - a_i), \epsilon(a_j - a_i)] \quad (35)$$

$$\Delta_2 \sim N[0, \sigma(a_j - a_i)] \quad (36)$$

Figure 13 shows the sensitivity of (a) CV^2 and (b) mean lost time holding to the error terms, ϵ and σ . The methods recommended in Bartholdi and Eisenstein (2011) and Daganzo and Pilachowski (2011) are shown with $\sigma = 0$ because they only require $E[A_j]$. The method proposed in Berrebi et al. (2015) is shown with

$\sigma = \{0, 0.2, 0.3, 0.4\}$ ¹⁸. For each value of ϵ , solid lines show the case where $\sigma = 0$, i.e. all trajectories equal $a_j + \Delta_1$. These lines describe the simulated performance of each holding method that would result if transit agencies used expected arrival times instead of probability distributions. The dashed lines describe the outcome of considering uncertainty, σ , surrounding the expected synthetic trajectories, which contain error distributed with accuracy parameter, ϵ .

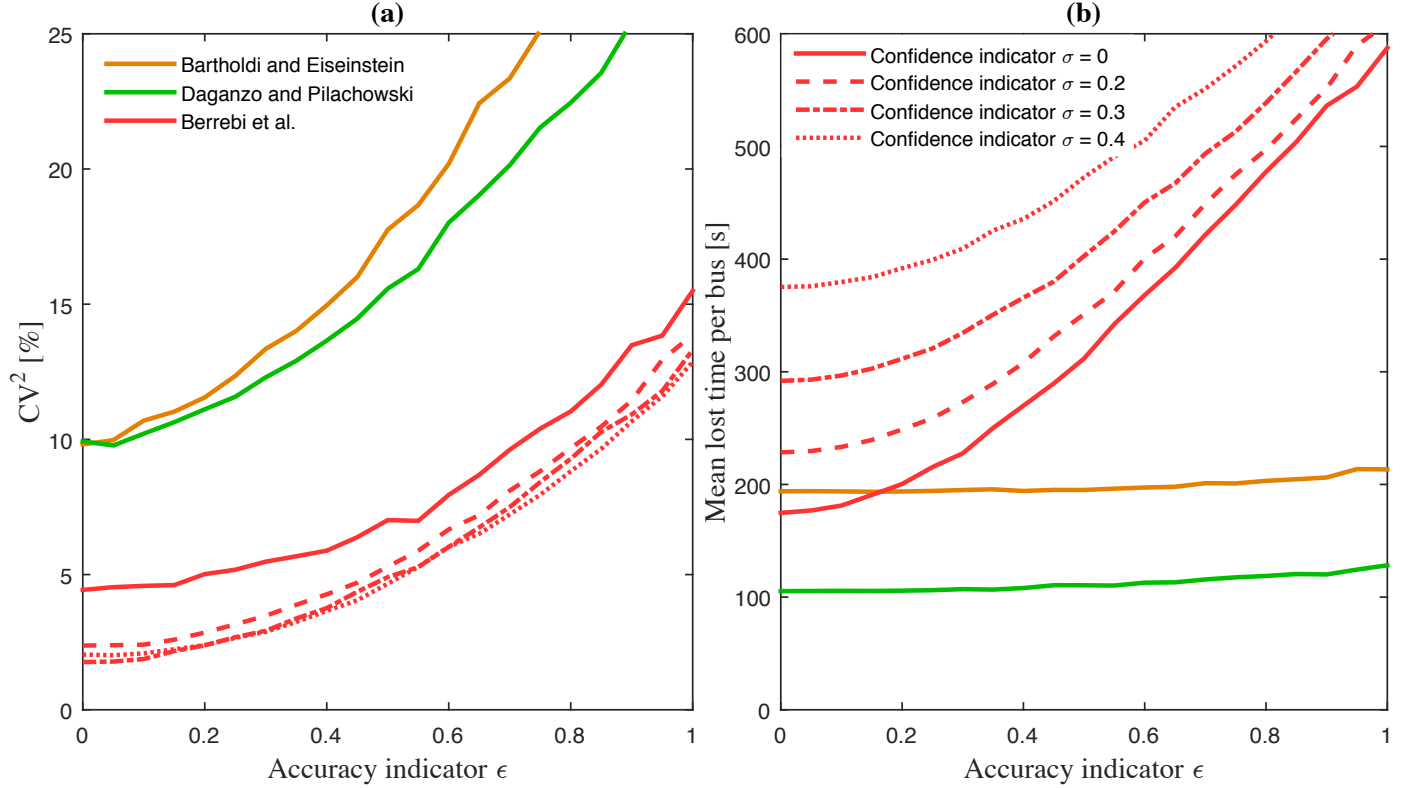


Figure 13: Sensitivity of (a) CV^2 and (b) Mean holding time to the maximal error in prediction

The methods recommended in Daganzo and Pilachowski (2011) and in Bartholdi and Eisenstein (2011) are affected by prediction accuracy in similar ways. When ϵ is null (perfect predictions), they dispatch buses with almost exactly the same

¹⁸We did not include $\sigma = 0.1$ because it closely overlapped with $\sigma = 0$.

CV^2 . As ϵ increases, they both destabilize at a fast rate (although CV^2 grows at a slightly greater rate for the method in Bartholdi and Eisenstein (2011)). The error in accuracy, ϵ , causes these methods to dispatch buses too soon sometimes and too late other times. The method in Daganzo and Pilachowski (2011) can dispatch buses with roughly half of the mean lost time in Bartholdi and Eisenstein (2011). The mean lost time is only slightly affected by increasing ϵ because the methods consider the expected arrival time, whose error is centered around the true mean.

For every value of ϵ and σ , the method in Berrebi et al. (2015) can dispatch buses with less than half the CV^2 of the other two methods, but it requires more holding time as ϵ and σ increase. The method in Berrebi et al. (2015) can dispatch buses with slightly more stable headways when it considers the uncertainty around its prediction, σ , but considering more uncertainty requires much more holding time. The reason for this increase in lost time is that the expected maximum of random variables is a monotonous increasing function of their variability. The confidence σ also causes holding time, especially when ϵ is small, for the same reasons. These results indicate that it may be adequate to replace the joint probability distribution of bus arrival times by their expectations to sacrifice headway stability for holding time and computational simplicity.

3.8 Conclusion

In this paper, we compared the performance of closed-form bus holding methods used in practice and recommended in the literature on Tri-Met route 72 in Portland, Oregon. We applied control at one and several control points along the route and tested how each method holds vehicles to stabilize headways, reduce passenger waiting time, and prevent bus bunching. We used a new prediction tool developed in Hans et al. (2015) to simulate the performance of real-time holding methods, which was essential to apply several holding methods studied. In addition, we coupled the

methods in Bartholdi and Eisenstein (2011) and Berrebi et al. (2015) with the method in Daganzo and Pilachowski (2011) to produce hybrid methods that can be applied at several control points along the route.

In the simulation, we found the trade-offs between headway stability and holding time for each holding method. The Schedule-Based methods can reduce the need for holding time but they are unable to stabilize headways. The Headway-Based methods can be parametrized and applied at several holding points to yield a wide range of holding times. The methods, however, are not competitive in terms of headway stability for any level of holding time. We found that Prediction-Based holding methods coupled with the prediction method in Hans et al. (2015) could produce competitive results in realistic settings. In particular, a hybrid between the method from Berrebi et al. (2015) and Daganzo and Pilachowski (2011), applied at one or several holding points, achieved the best compromises between headway regularity and holding time on a wide range of desired trade-offs.

When evaluating the impact of prediction accuracy on the performance of holding mechanisms, we found that prediction errors increased headway instability of real-time holding methods. Inaccurate predictions substantially increase mean lost time holding for the method in Berrebi et al. (2015). Using high quality predictions such as those in Section 3.5 for real-time holding methods is necessary to maintain stable headways and short holding times. The uncertainty considered in the prediction for the method in Berrebi et al. (2015) helps reduce CV^2 at a high cost of holding time, especially for highly accurate predictions. Replacing the probability distribution of expected arrival times by their mean could help reduce the mean lost time holding.

This paper compares holding methods assuming Poisson distributed arrivals, but this assumption is not always true (Luethi et al., 2006). The performance metrics do not reflect the value that schedules offer for both passengers and operators. In addition, on certain routes, schedule or real-time information coordinated arrivals

may impact the performance of holding methods compared. Future research should compare holding methods based on different passenger arrival distributions.

In this paper, TriMet Route 72 is used as a test bed for holding methods used in practice and recommended in the literature. There are routes resembling Tri-Met Route 72 in terms of passenger demand, traffic congestion, and land-use in almost every metro area in the United States. Route 72 is a typical high-frequency route (7-8 minute headways in peak hours) that faces the issue of bus bunching. We considered Route 72 as generally as possible, often applying sensitivity analyses. The analysis presented in this paper can therefore provide a basis for transit agencies to decide on a closed-form holding method on their high frequency routes. Every holding method presented here strikes a different balance between the conflicting objectives to stabilize headways and dispatch buses with little holding time.

As transit agencies look to implement innovative holding methods, further testing and simulation should be done for cases of unique passenger loading rules, severe passenger overflow, and other route characteristics that substantially deviate from TriMet Route 72. Future research should explore how route characteristics such as travel patterns, route instability, and perception of waiting time affect the desired trade-offs between these conflicting objectives. Finally future research should test and compare holding methods on a real bus route, and include optimization-based holding methods in the analysis.

CHAPTER IV

IMPLEMENTING A DYNAMIC HOLDING METHOD ON THREE HIGH-FREQUENCY TRANSIT ROUTES

4.1 *Abstract*

On high-frequency routes, buses tend to bunch together, creating gaps in service and causing undue passenger waiting time. There are many approaches to solving the bus bunching problem in the literature but there lacks empirical analysis on practical implementation. In this study, the proposed holding method is implemented on three high-frequency transit routes, the Atlanta Streetcar and the Georgia Tech Red Route in Atlanta, GA, and the VIA Route 100 in San Antonio, TX. The performance of the method is evaluated in terms of headway stability and holding time. The method is found to improve headway stability compared to the schedule on all three case studies, but requires longer holds in some cases. The impact of location data quality, prediction accuracy, the human element and the surrounding environment are evaluated. The main challenges to implementing real-time control are discussed and strategies to address them are recommended.

4.2 *Introduction*

High-frequency transit routes are inherently unstable. Passengers tend to arrive at stops without using a schedule (Fan and Machemehl, 2009). At each stop, the number of passengers waiting for a bus is proportional to the time since the last vehicle left; as the headway of a bus grows, so does the number of passengers boarding and alighting at each stop and *vice versa*. Boarding these passengers further delays lagging vehicles. Long headways tend to get longer and short headways tend to get shorter. Eventually,

vehicles bunch together and travel as a platoon, creating large gaps in service that cause undue passenger waiting time.

Although passengers on high-frequency routes tend to disregard the schedule, time-tabling vehicle departures from time-points¹ along the route is still the method used by almost all transit agencies. To allow lagging vehicles to recover, planners include buffer time in their schedules as high-percentiles of historical running time (Furth et al., 2006). As a result, vehicles systematically waste time holding at control points when operating conditions are fluid, and consistently miss their scheduled departure times when operating conditions are congested. The schedule is therefore not useful for passengers who arrive randomly at stops nor effective for transit agencies who strive to stabilize headways.

Since the early 1970’s researchers have developed innovative methods to stabilize headways on high-frequency routes using real-time information (Osuana and Newell, 1972; Barnett, 1974; Newell, 1974). It was not until recently, however, that vehicle location data collection could be automated, which made real-time holding methods feasible. A rapidly growing body of literature, which is reviewed in Berrebi et al. (2015), has emerged since the early 2000’s. Berrebi et al. (2016) compared closed-form holding methods used in practice and recommended in the literature based on data from TriMet Route 72 in Portland, OR. The study found that the holding method in Berrebi et al. (2015) can dispatch vehicles with more stable headways than other closed-form methods used in practice and recommended in the literature with the same mean holding time.

The method in Berrebi et al. (2015) takes a global approach to the bus bunching problem. The method probabilistically identifies the vehicle with the most delay on the route, and holds each preceding vehicle to diffuse large gaps in service. The method uses the probability distribution of vehicle arrival times as input. Although

¹Time-points are control points where vehicles wait for their scheduled departure time

there are prediction methods in the literature that can generate distributions, such as Hans et al. (2015), it is simpler to generate predictions deterministically as expected arrival times. Therefore, the first step to implement the holding method proposed in Berrebi et al. (2015) is to derive its deterministic equivalent.

Transit agencies who wish to use the method recommended in Berrebi et al. (2015) or any other method to stabilize headways on their high-frequency routes need insight on the performance of the method in realistic settings, and the main challenges to implementation. Berrebi et al. (2016) introduced sensitivity analyses to evaluate the impact of parametrization, prediction accuracy, and control point density on closed-form holding methods using the most realistic framework possible in a simulation. As with all simulation experiments, however, the study is limited to a theoretical interpretation of the transit route. The simulation cannot consider all the small perturbations that tend to accumulate and lead to the destabilization of a route. Therefore, it may not fully represent the compounded effects of holding methods on transit operations. In particular, there are several domains, in which the simulation experiment may diverge from a real application:

- In the simulation experiment, vehicle location data is assumed to be always accurate and always available. In reality, however, the GPS signal wanders and sometimes even gets lost for minutes at a time. In addition, vehicle location data is made available at a set frequency, thereby creating a lag between the time of recording and the time when the data becomes available. The accuracy and availability of vehicle location data has the potential to affect the implementation of real-time holding methods.
- The simulation found that when the standard deviation of prediction error was less than ten percent of the horizon (time between prediction made and vehicle arrival), prediction accuracy had little effect on the performance of the method

in Berrebi et al. (2015) and others. If the dwell time at the control point is not considered, however, vehicles that need to be dispatched immediately may be delayed. This delay can undermine the ability of the holding method to maintain stable headways.

- In a simulation experiment, operators can perfectly follow holding instructions. In reality, however, transit vehicles are operated by human beings, whose perception, cognition, and behavior affect operations. Operators may not adhere to unclear or impractical instructions, which may disrupt the route. They may also need to take bathroom breaks, which can delay their departure time. On the other hand, operators can use their intuition, which is informed by unique live information and experience, to amplify the impact of the holding method.
- Transit vehicles that run in mixed traffic are particularly prone to bus bunching because they have to compete for capacity and are subject to random disruptions. Some of these disruptions are periodically recurring, others are discrete events; some are localized, others span the entire route. Although these disruptions can affect the capacity of a holding method to stabilize a high-frequency transit route, they often cannot be modeled explicitly in a simulation.

To evaluate the performance of the holding method in realistic settings and assess challenges to implementation, a deterministic version of the method recommended in Berrebi et al. (2015) is derived analytically. The method is applied on three high-frequency transit routes: the Atlanta Streetcar and the Georgia Tech Red Stinger Route in Atlanta, GA, and the VIA Route 100 in San Antonio, TX. The performance of the method is compared to the schedule that was currently in use. The impact of data and prediction quality on the performance of the method and ease of implementation are assessed. In addition, the level of adoption by supervisors, dispatchers and operators and their effect on compliance are analyzed. The effect of the route

environment on the implementation is also evaluated. Finally, the study discusses the lessons learned from implementing a real-time dispatching method in high-frequency transit systems.

4.3 Literature Review

Although the literature on holding methods to avoid bus bunching is abundant, there are few studies on the implementation of these methods in live experiments. Abkowitz and Engelstein were the first to present a solution to the bus bunching problem that could be implemented in a real route (Abkowitz and Engelstein, 1984). The paper introduced a closed form holding method that uses real-time vehicle location data. The method was parametrized off-line by mathematical programming using historical data. Although the authors did not report a live implementation, the dispatching software and a path to implementation were described. The protocol would communicate holding instruction directly to operators.

Pangilinan et al. compared the Naive Headway method with the method from Turnquist (1982) on CTA Route 20 in Chicago, IL (Pangilinan et al., 2008). The experiment was implemented during the morning peak hour for four consecutive days. Supervisors were located at four control points along the route, with a dispatcher instructing them by phone to hold vehicles. The study found that the method from Turnquist (1982) yielded less headway variability than the Naive Headway. However, when comparing the results with a simulated version of the same experiment, the authors found that the simulation yielded much more stable headways. The authors reported that the main difficulty stemmed from the dispatcher’s capacity to manually compute holding times at four control points simultaneously.

Argote et al. extended the method from Xuan et al. (2011) to overlapping bus routes (Argote-Cabanero et al., 2015b). The method was tested in San Sebastian in

Spain, along a busy corridor where two unsynchronized bus routes overlap. Cruising guidance was provided to operators through tablets. The method was able to improve schedule adherence compared to a Naive Schedule method. The authors also tested the compliance of operators to instructions provided by the tablets. They found that parametrizing their method to yield shorter holding times increased driver compliance.

Lizana et al. implemented the method recommended in Delgado et al. (2012) on Transantiago Route 210 in Santiago de Chile (Lizana et al., 2014). Instructions to hold, slow down, and increase speed were communicated to the operators along the route via tablets connect by 3G. The study found that the method was able to reduce bus bunching compared to the schedule. The authors noted that the main challenges to implementation were technology failure and operator compliance.

The implementations in the literature provide valuable insights to transit agencies who wish to implement not only the methods tested in the studies but any method that uses real-time vehicle location data to inform holding decisions. When making the transition from simulation to live implementation, holding methods are confronted with a plethora of practical challenges. In particular, the data quality, prediction accuracy, human factors, and route characteristics play an import part in the success of real-time holding methods. These factors have been insufficiently studied. Transit agencies need the analysis to support the widespread implementation of real-time holding methods

4.4 Deriving Deterministic Method

In the simulation experiment, predicted vehicle arrivals were generated for the method in Berrebi et al. (2015) as probability distributions. Berrebi et al. (2016) found that the distributions could be replaced deterministically by the expected vehicle arrival times as discrete distributions with probability equal to one. The deterministic

version of the method yielded more unstable headways but requires less holding time overall. It remains, however, that the formula for the holding method in Berrebi et al. (2016) is more complicated than needed. This section derives a simpler expression for the holding time recommended by the method in Berrebi et al. (2015) when using deterministic predictions instead of probability distributions.

Consider that bus i has just arrived at the control point at current time a_i . All downstream buses are indexed in ascending order. We know that the last vehicle to depart the control point was vehicle $i - 1$ at time d_{i-1} . We must decide how long to holding vehicle i , based on the prediction that vehicles $i + 1, i + 2, \dots$ will arrive at times A_{i+1}, A_{i+2}, \dots . In Berrebi et al. (2015), A_{i+1}, A_{i+2}, \dots are random variables with known probability distributions. Lemma 4.4.1 asserts that in the special case when A_{i+1}, A_{i+2}, \dots are point estimates, the holding method in Berrebi et al. (2015) is equivalent to the difference between vehicle i 's backward headway and the longest possible average headway on the route.

Lemma 4.4.1. *The holding method in Berrebi et al. (2015) can be formulated as*

$$\max_{r-i} \frac{A_r - d_{i-1}}{r - (i-1)} - (a_i - d_{i-1}) \text{ when } A_j \text{ is a discrete value for all } 0 \leq j \leq n$$

Proof.

$$\frac{E \left[\max_{r-i} \frac{A_r - a_i}{r-i} - (a_i - d_{i-1}) \right]}{1 + E \left[\left(\arg \max_{r-i} \frac{A_r - a_i}{r-i} \right)^{-1} \right]} \quad (37)$$

$$= \left[\max_{r-i} \frac{A_r - a_i}{r-i} - (a_i - d_{i-1}) \right] \frac{1}{1 + \left(\arg \max_{r-i} \frac{A_r - a_i}{r-i} \right)^{-1}} \quad (38)$$

$$= \left[\max_{r-i} \frac{A_r - a_i}{r-i} - (a_i - d_{i-1}) \right] \frac{\left(\arg \max_{r-i} \frac{A_r - a_i}{r-i} \right)}{1 + \left(\arg \max_{r-i} \frac{A_r - a_i}{r-i} \right)} \quad (39)$$

$$= \left[\max_{r-i} \frac{A_r - a_i}{r-i} \right] \frac{\left(\arg \max_{r-i} \frac{A_r - a_i}{r-i} \right)}{1 + \left(\arg \max_{r-i} \frac{A_r - a_i}{r-i} \right)} - (a_i - d_{i-1}) \frac{\left(\arg \max_{r-i} \frac{A_r - a_i}{r-i} \right)}{1 + \left(\arg \max_{r-i} \frac{A_r - a_i}{r-i} \right)} \quad (40)$$

We then notice that because $\{A_2, A_3, \dots\}$ are deterministic, the denominator in $\max_{r-i} \frac{A_r - a_i}{r-i}$ is equal to $\arg \max_{r-i} \frac{A_r - a_i}{r-i}$. To simplify the first expression in the rhs of (4), we define r^* , for which:

$$r^* - i = \arg \max_{r-i} \frac{A_r - a_i}{r-i} \quad (41)$$

Equation (4) can now be simplified to:

$$\frac{A_{r^*} - a_i}{1 + \left(\arg \max_{r-i} \frac{A_r - a_i}{r-i} \right)} - (a_i - d_{i-1}) \frac{\left(\arg \max_{r-i} \frac{A_r - a_i}{r-i} \right)}{1 + \left(\arg \max_{r-i} \frac{A_r - a_i}{r-i} \right)} \quad (42)$$

$$= \frac{A_{r^*} - a_i}{1 + r^* - i} - (a_i - d_{i-1}) \frac{r^* - i}{1 + r^* - i} \quad (43)$$

$$= \frac{A_{r^*} - a_i + (a_i - d_{i-1})}{1 + r^* - i} - (a_i - d_{i-1}) \frac{r^* - i}{r^* - i} \quad (44)$$

$$= \frac{A_{r^*} - d_{i-1}}{r^* - (i - 1)} - (a_i - d_{i-1}) \quad (45)$$

$$= \max_{r-i} \frac{A_r - d_{i-1}}{r - (i - 1)} - (a_i - d_{i-1}) \quad (46)$$

The expression is much more simple than Equation (1). It can be computed manually by taking the difference between the greatest relative frequency of arrival ($\max_{r-i} \frac{A_r - d_{i-1}}{r - (i-1)}$) and the backward headway of the holding vehicle ($a_i - d_{i-1}$). It now makes intuitive sense that the DHA is consistent with Figure 2 of Berrebi et al. (2015)

4.5 *Case Study*

4.5.1 **DynamicTime**

For the purpose of this research, we developed an open-source dispatching software: DynamicTime. DynamicTime is a web application accessible through any device with an internet browser and web connection. DynamicTime is an extension of the open-source tool TransiTime. TransiTime is a software program that uses vehicle location data to predict vehicle arrivals. For this implementation, TransiTime was modified to generate predictions on frequency-based service without the need for a schedule. The prediction method simply averages historical travel time on each route segment by time of day. The predictions are then used to compute holding times recommended by the holding method derived in Lemma 4.4.1. Finally, the holding times are displayed in a dashboard.

Figure 14 shows a screen capture of the DynamicTime dashboard. Each time a vehicle arrives at the control point, the DynamicTime dashboard starts showing a green countdown and makes a sound to notify the dispatcher. When the countdown reaches zero, another sound is emitted, the font turns red, and start counting up until the vehicle leaves the control point. The dashboard provides information about the vehicle currently at the stop and the next vehicle expected to arrive. A map also shows the current location of each vehicle on the route. Dispatchers can click on individual stops or vehicles to obtain additional information and predictions.

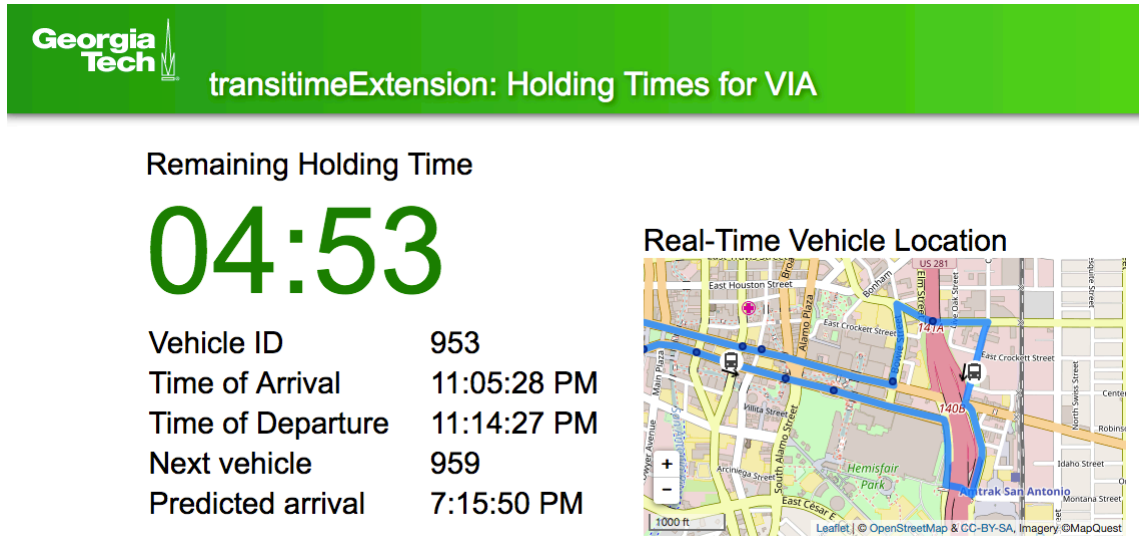


Figure 14: Screen capture of the DynamicTime dashboard

4.5.2 Transit Systems

The holding method was implemented on three high-frequency transit routes: the Atlanta Streetcar, the VIA Primo 100 Route, and the Georgia Tech Red Stinger Route. These three routes offer widely varied test cases. The first is a streetcar route running in the heart of Downtown Atlanta, the second is a Bus Rapid Transit route connecting San Antonio suburbs to downtown through Highway I-10, and the third serves a student population that surges before and after class. Each route runs in mixed traffic on a schedule that is unavailable to the public.

4.5.2.1 Atlanta Streetcar

The Atlanta Streetcar is a 2.7 mile transit line operated by the City of Atlanta, GA. Figure 15 shows a map of the Atlanta Streetcar route. Streetcars run in mixed traffic in the Downtown and Sweet Auburn neighborhoods. The route is subject to heavy traffic from surrounding land-uses and entrances to the I-75/85 interstate. Passenger demand fluctuates unpredictably throughout the day, coming from Georgia State University, and other sports and event centers located nearby. The Peachtree Center

heavy rail station also brings passengers to the Streetcar from Metropolitan Atlanta Rapid Transit Authority (MARTA).

The Atlanta Streetcar usually runs two vehicles based on a schedule with 15 minute headways. For the experiment, the City of Atlanta agreed to run three vehicles simultaneously. The vehicles ran on a schedule with 10 minute headways for five days to provide historical data, which was used to inform predictions for the implementation. Bus bunching was not a problem on the Streetcar route with either two or three vehicles. The 2-vehicle schedule has three time points along the route: King Historic District at the eastern end, Woodruff Park in the middle, and Centennial Olympic Park at the western end of the route. For the implementation of the method, we used a single control point at Centennial Olympic Park (red star), which is the only stop where vehicles have a dedicated space to hold.

The proposed holding method was implemented between 11 AM and 2 PM for eight weekdays. Only four days yielded usable data due to traffic collisions and construction projects along the route. Dispatchers inside the Vehicle Maintenance Facility were receiving holding instructions from the DynamicTime dashboard displayed on their desktop monitor. Each time a vehicle arrived at the control point, dispatchers communicated holding instructions to the operators by radio. In the case of severe disruptions, such as a traffic collision, dispatchers had to apply control at intermediate time points (King Historic District and Woodruff Park) to prevent Streetcars from getting too close, which would damage the electrical circuits. Holding instructions at intermediate control points were based on dispatchers' estimated headways based on the real-time map of vehicle locations provided by DynamicTime.

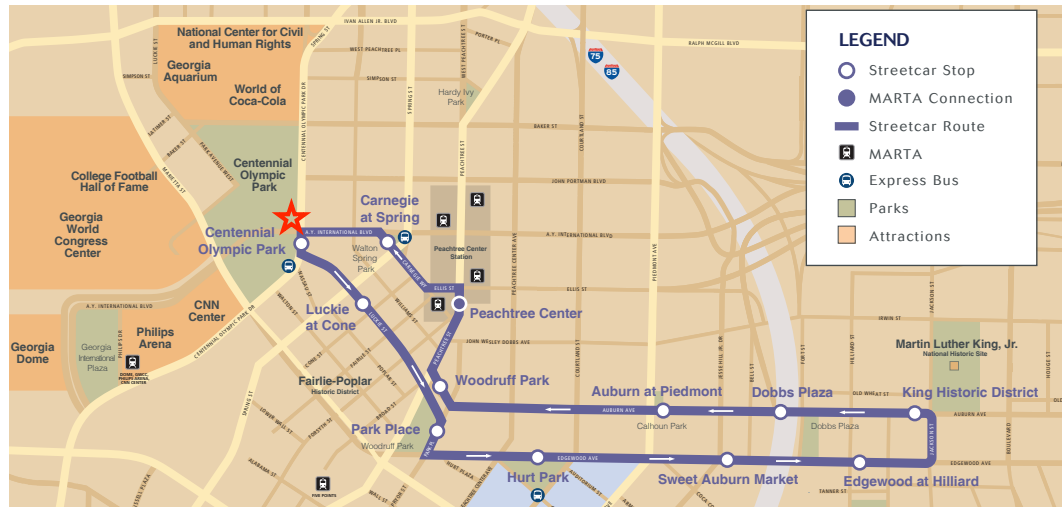


Figure 15: Map of the Atlanta Streetcar route adapted from Atlanta (2016)

4.5.2.2 *VIA Primo*

The VIA Route 100 is a 20-mile Arterial Rapid Transit line connecting the San Antonio Central Business District to the South Texas Medical Center as shown in Figure 16. Although vehicles are not running on dedicated right-of-way, intersections are equipped with Traffic Signal Priority capabilities. Articulated buses carry 6,000 passengers per day and serve 16 stops, which have waiting platforms and real-time vehicle arrival displays. From 8 AM to 6 PM during weekdays, eleven buses run simultaneously at 10-minute headways according to a schedule that is unavailable to the public. The route has a severe bus bunching problem. The proposed holding method was successfully implemented between 2:30 and 5:30 PM Central on Tuesday, May 9, and on Friday, May 12.

Although there are four time-points used for schedule recovery, only two have dedicated space, where vehicles can safely hold for long periods of time without impeding surrounding traffic. For the implementation, these two stops were selected as control points: South Texas Medical Center at the northern end and Ellis Alley Park

and Ride at the southern end (red stars). A supervisor was positioned at each control point. Supervisors were equipped with tablets that displayed the DynamicTime dashboard using a WIFI connection. Each time a vehicle arrived at the control point, supervisors informed operators how long they should hold. Supervisors had a direct line of communication with their management and with the research team by radio. There were several instances during the implementation when DynamicTime did not display holding times correctly due to a defect in the software. The research team had to calculate holding times manually and communicate them to supervisors. The process did not disrupt the experiment.



Figure 16: Map of VIA Route 100 (VIA, 2017)

4.5.2.3 Georgia Tech Stinger

The Red Stinger Route is a 2.5 mile loop serving students, faculty and staff on Georgia Tech campus. Figure 17 shows a map of the system. Four vehicles run clockwise every six minutes from 7 A.M. to 5:30 P.M. during weekdays (the Blue Stinger Route follows the same path but runs counter-clockwise). Buses stop at two time-points along the route to recover their schedule, which is unpublished. Passenger demand tends to surge at the beginning and end of classes, causing crowding. The route is also subject to traffic congestion. As a result, bus bunching is a common occurrence on the route.

On this route, the GPS devices in the vehicles have a tendency to lose the signal for

extended periods of time. The research team attempted to implement the proposed method four times but had to abandon due to unresponsive GPS units. On the fifth attempt, the holding method was successfully implemented between 3:13 and 4:30 PM. During the implementation, members of the research team communicated holding instructions directly to operators at the control point, which was located at the intersection between Ferst Drive and Atlantic Drive (red star). The control point had dedicated space for vehicles to hold without impeding traffic. The implementation ended when one operator ran out of gas and had to go re-fuel at the maintenance yard.

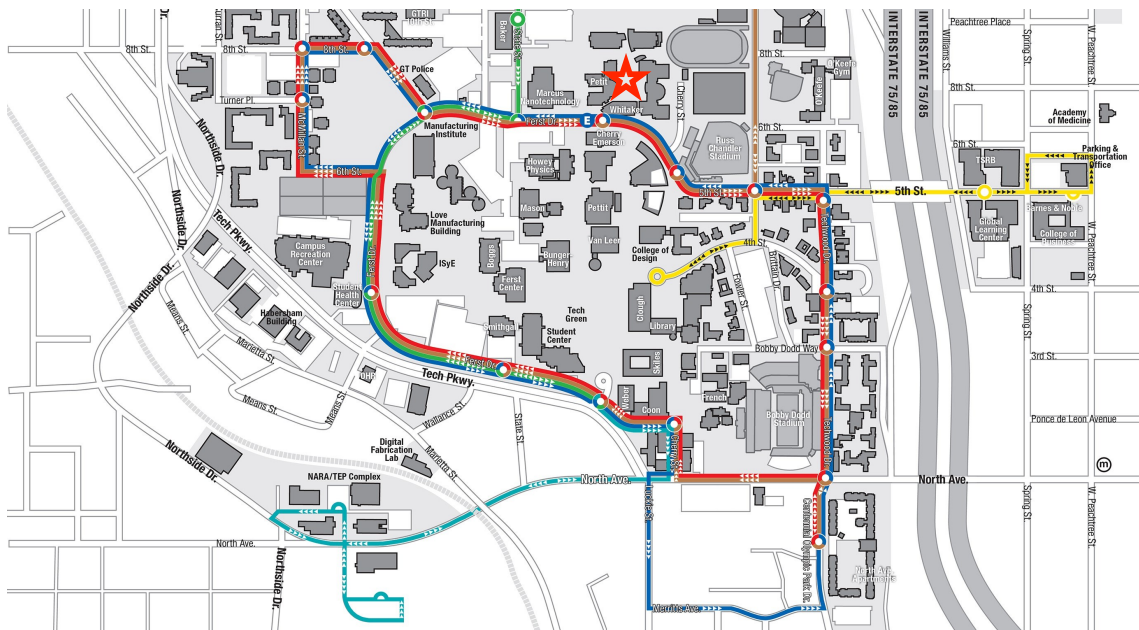


Figure 17: Map of Georgia Tech Red Stinger Route (Parking and Services, 2017)

4.6 Results

4.6.1 Atlanta Streetcar

Although the Atlanta Streetcar usually only runs two vehicles at a time, the City of Atlanta agreed to run three vehicles in early February to provide the research team with historical data. The data were used to train the TransiTime prediction algorithm. The proposed method was then successfully implemented for two days in

March, one in April, and one in May. This subsection reports our findings based on four weekdays of the 3-vehicle schedule and four weekdays of the proposed method. All other data collected were invalidated by traffic collision and construction projects impeding the right of way. We first analyze the results on a typical implementation day, then compare headways and holding times for the proposed method with the three-vehicle schedule.

The headways at departure from the control point (black), the actual holding times (gold), and holding times recommended by the DynamicTime application (blue) are shown in Figure 18 for May 23, 2017. The scope of the figure is limited to the implementation period because the route was operated by only two vehicles before and after the implementation. During the study period, vehicles were dispatched with relatively stable headways with some fluctuation in the first hour and a half and much less in the second part. Most of the fluctuation was caused by the discrepancy between recommended and actual holds.

The low adherence to instructions was overwhelmingly caused by the actuated traffic signal located at the control point. Operators often missed their permissive phase and had to wait for the entire 120-second cycle. On average vehicles held for 107 seconds longer than recommended by the DynamicTime dashboard. Adherence to instructions could be improved by providing holding instructions directly to operators via tablets in the vehicles or kiosks at stops and synchronizing the method with the traffic cycle. Reducing missed cycles may improve both headway reliability and reduce holding times.

To compare operations under the schedule with the proposed method, we have analyzed headways and holding times as histograms. Figure 19 shows a histogram of headways at departure from the control point as dispatched by the schedule (white) and by the proposed method (black)². The schedule method dispatched vehicles

²We chose to focus the histogram on the range of data instead of starting the abscissa at zero.

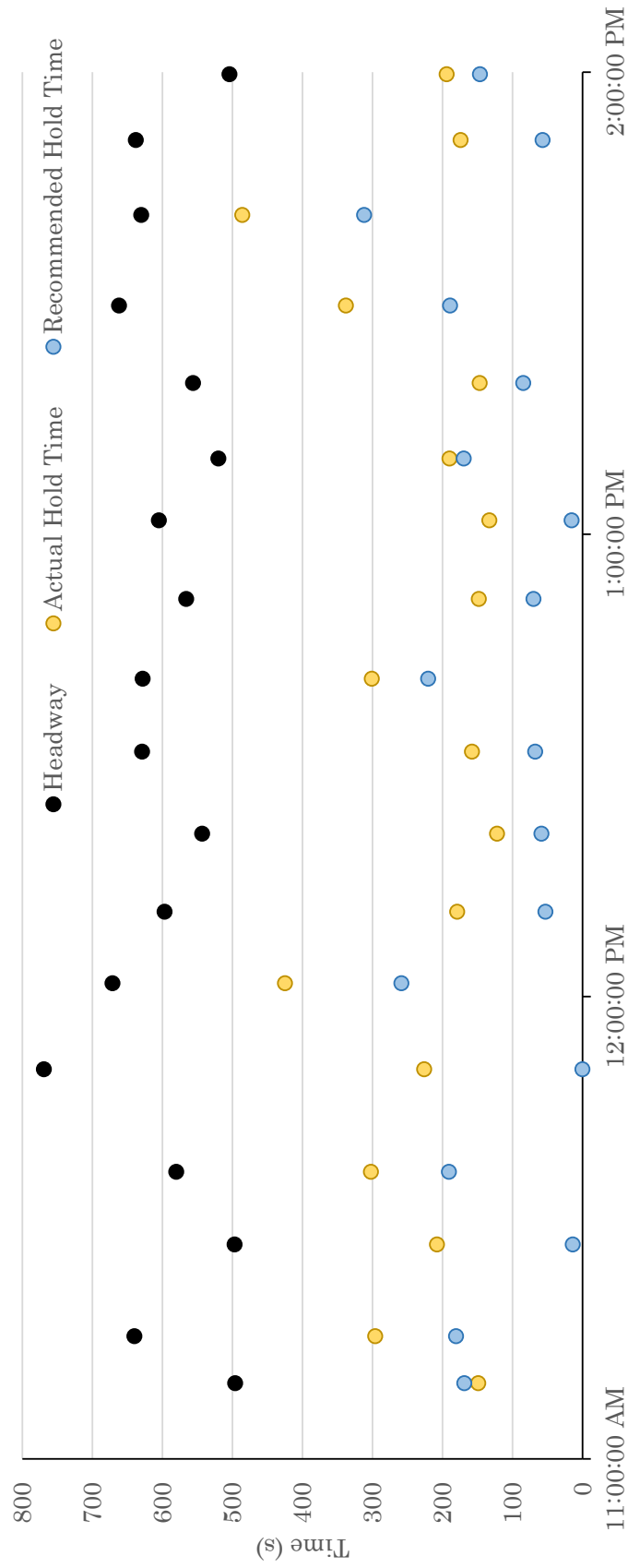


Figure 18: Headways at departure from the control point (black), actual holding times (gold), and holding times recommended by the DynamicTime application (blue) as a function of time.

with a wide range of headways; 39% of headways were shorter than 480 seconds and 12% were greater than 780 seconds. Under the proposed method, the distribution of headways was more compact with 60% of vehicles dispatched with headways between 540 and 660 seconds. The proposed method, however, dispatched buses with slightly longer headways, 622 seconds on average, compared to 584 seconds for the schedule method.

Finally, we compared holding times under the schedule with the proposed method. Figure 20 shows a histogram of the actual holding times in seconds for the schedule method (white) and the proposed method (gold), and the recommended holding times by the proposed method (blue). The schedule method was able to keep holding times at only 125 seconds on average. The proposed method required 185 seconds of holding time on average with a wider spread. The holding time recommended by the proposed method is substantially shorter than actual holds under the method.

4.6.2 San Antonio VIA

The proposed method was successfully implemented on VIA Route 100 for two days. Because the operating conditions were substantially different, we present the days separately as two case studies.

4.6.2.1 Tuesday, May 9

On Tuesday, May 12, the proposed holding method was able to maintain stable operations. Figure 21 shows headways and holding times at the northern (top) and southern (bottom) control points ³. Before the start of the implementation, headways of departing vehicles (hollow points) were relatively stable at the northern control point and rather unstable at the southern control point, where they oscillated between 300 and 900 seconds. At 2:30 PM, the DynamicTime application started instructing supervisors to hold vehicles. Headways (black points) maintained the same level of

³The figure only goes until 6 PM because VIA started taking vehicles off the route at that time

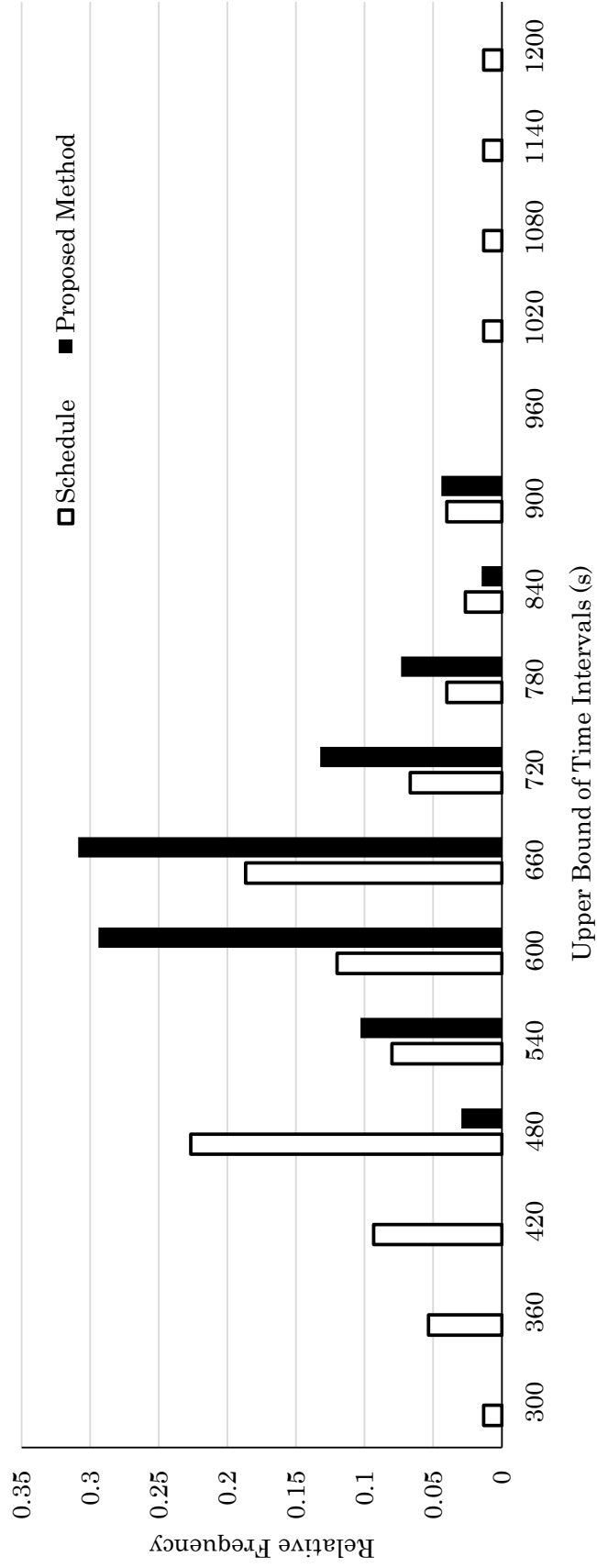


Figure 19: Histogram of headways of at departure from the control point as dispatched by the schedule (white) and by the proposed method (black).

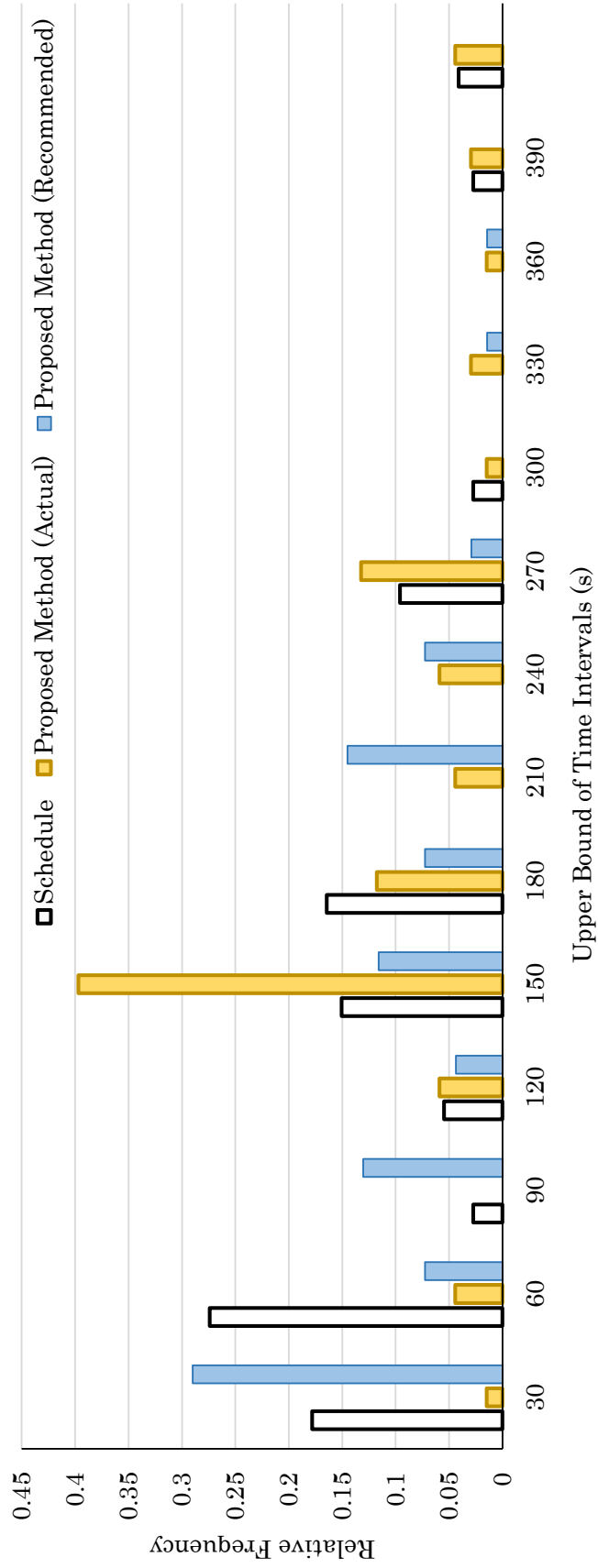


Figure 20: Histogram the actual holding times in seconds for the schedule method (white) and the proposed method (gold), and the recommended holding times by the proposed method (blue).

stability at the northern control point as they had in mid-day off-peak, with the exception of one vehicle that left with a headway of 803 seconds at 3:13 PM. The vehicle waited for longer than recommended by the method due a miscommunication between the research team and the supervisor. At the southern control point, vehicles were dispatched with greater stability than earlier in the day.

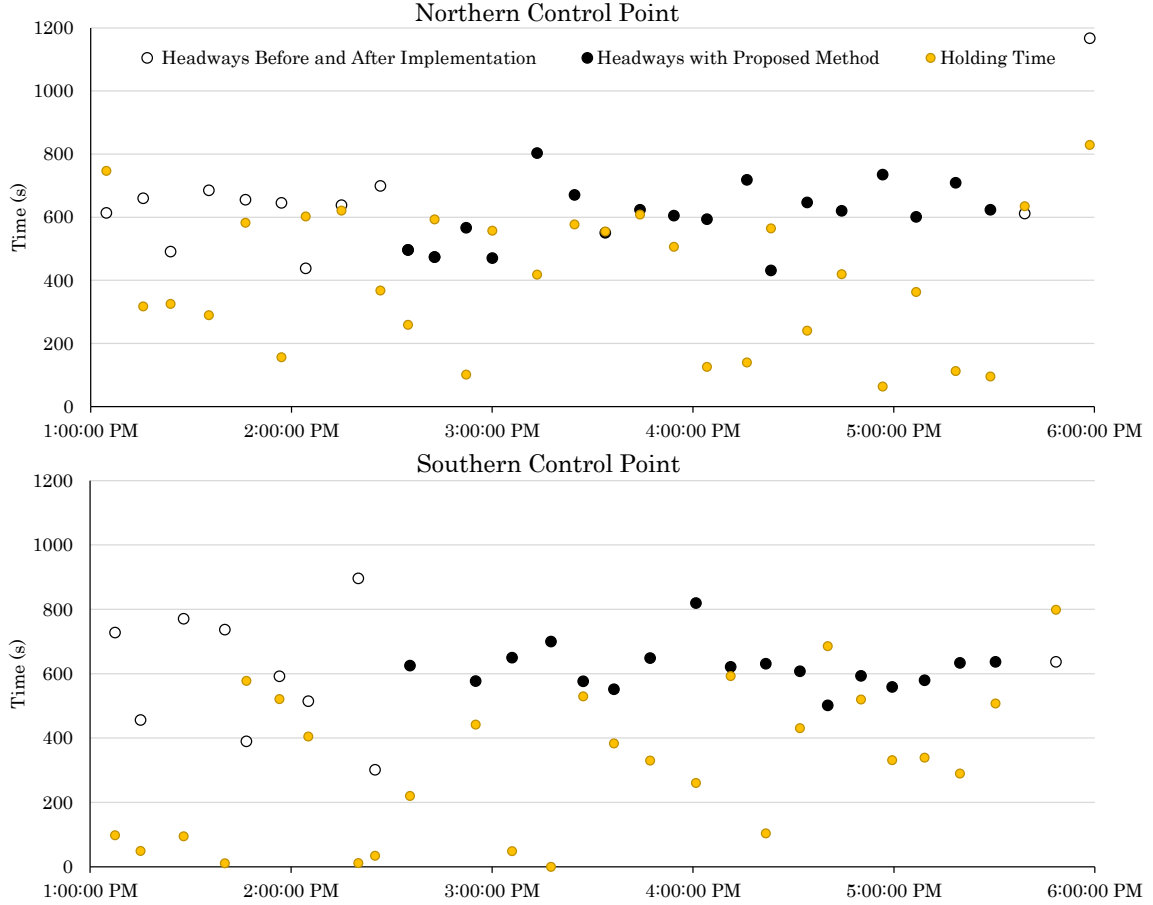


Figure 21: Headways and holding times from the proposed method at the northern (top) and southern (bottom) control points on Tuesday, May 9.

In order to evaluate the impact of the proposed method on operations, it is interesting to look at route-level stability. Figure 22 shows the headway Coefficient of Variation squared (CV^2) over the implementation period. The CV^2 is a measure of headway stability calculated as the ratio between the variance and mean headway

squared over every vehicle on the route at a given time. To reduce the noise, CV^2 is calculated as a simple rolling average with the five measurements before and after. The implementation day, May 9, is compared with Tuesdays January 3, 19, and 31, for the same time period. We used every day of historical data available that did not have missing vehicles or missing data.

On Figure 22, the days that started with high CV^2 remained disrupted and the days that started with a low CV^2 maintained stable headways. On January 17 and 31, the Route 100 was already severely bunched at 2:30 PM with a CV^2 close to 0.8. On these days, the schedule was unable to stabilize headways and CV^2 remained close to or above 0.5 until 5:30 PM. On January 3 and May 9, the route was relatively stable at 2:30 PM with CV^2 at approximately 0.2 for both methods. Although January 3 was not an official holiday, the traffic conditions and passenger demand were presumably much calmer than on May 9. Yet, starting at 4 PM, the route quickly destabilized under the schedule. On May 9, while the proposed method was being implemented, operations remained stable until 4 PM, when a large headway caused CV^2 to triple almost instantly. The proposed method, however, was able to recover stable operations. The proposed method yielded the lowest CV^2 for 2:24 hours out of the three hour period.

4.6.2.2 Friday, May 12

After the successful implementation of May 9, the proposed method was implemented again on May 12 with different results. Figure 23 shows the headways and holding times at the northern (top) and southern (bottom) control points. In the mid-day off-peak period, vehicles departed both control points with widely uneven headways. A few minutes before the start of the implementation, one vehicle left the northern control point with a 1012 second headway followed by another vehicle 124 seconds later. The proposed method started increasing headways from the southern control

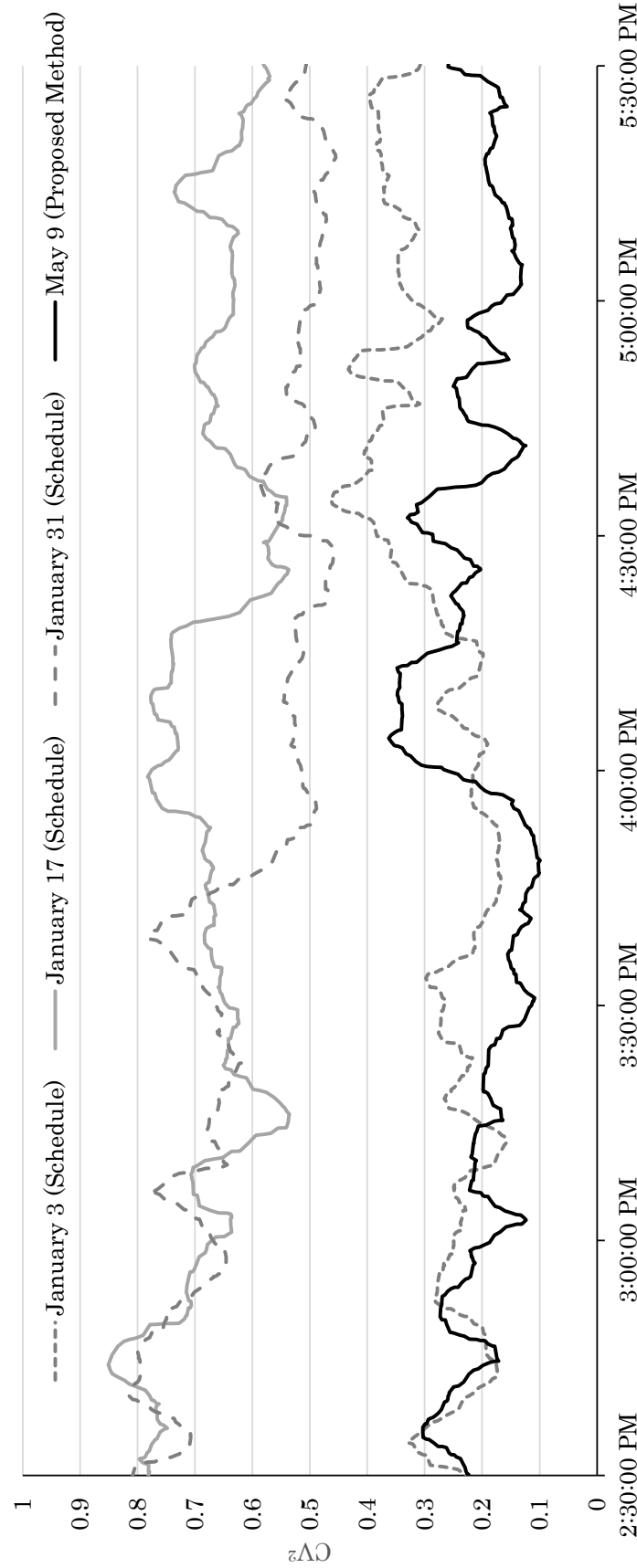


Figure 22: Headway Coefficient of Variation (CV^2) throughout the route between 2:30 and 5:30 under the schedule (January 3, 17, and 31) and under the proposed method (May 9)

point progressively in anticipation for the lagging bus. The vehicle left the southern control point with a headway of 986 seconds after unloading and boarding passengers for 163 seconds, and arrived at the northern control point with a headway of 1825 seconds. Since the beginning of the implementation, eight vehicles had been dispatched from the northern control point with headways within an 85 second range. The lagging vehicle held for an additional 348 seconds for a bathroom break, blocking the two following vehicles that were already bunched. When it departed with a headway of 1378 seconds, another bus was dispatched immediately due to miscommunication between the research team and the supervisor. Despite this incident, the proposed method was able to slowly stabilize the dispatching process for the remaining hour of implementation.

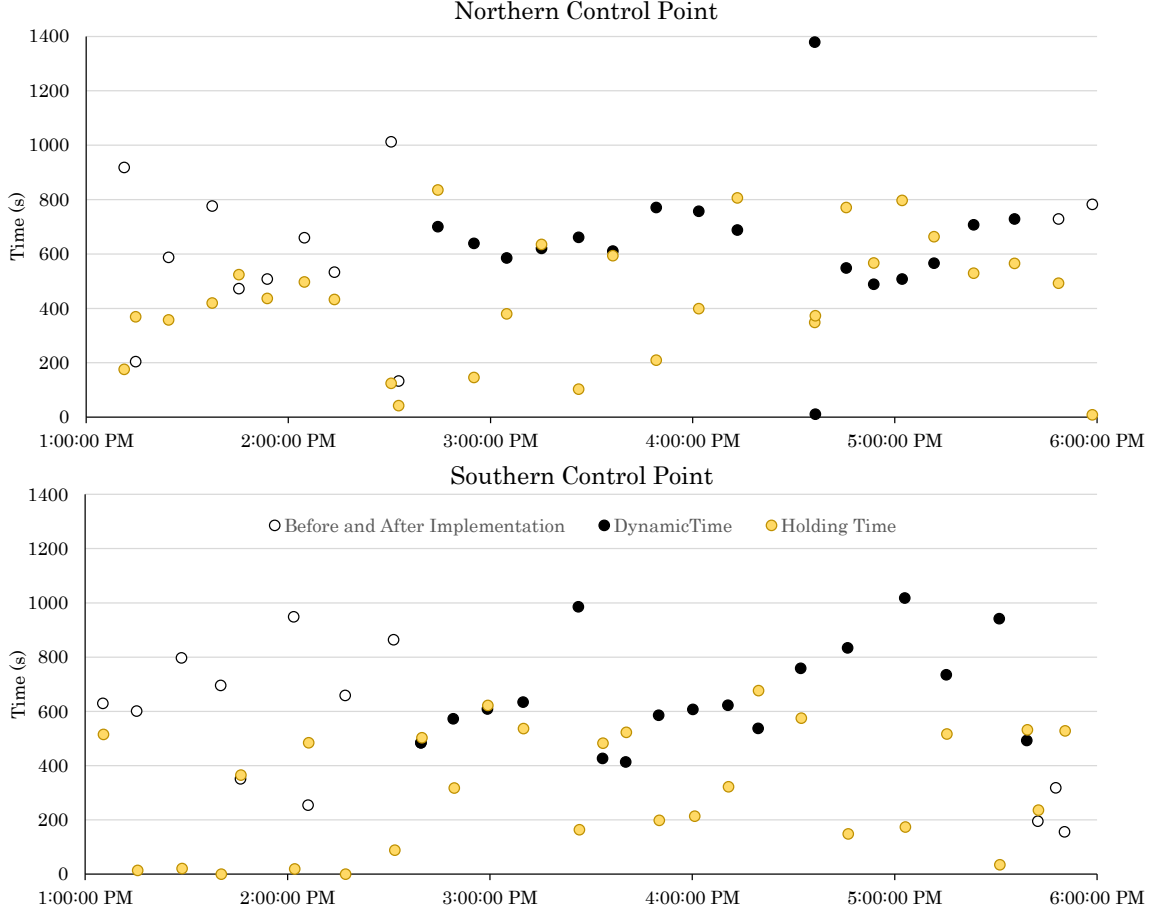


Figure 23: Headways and holding times from the proposed method at the northern (top) and southern (bottom) control points on Friday, May 12.

To evaluate the global headway stability during the implementation, we compared May 12 with Fridays January 13, and 20, when the route was controlled by the schedule. Figure 24 shows CV^2 from 2:30 to 5:50 PM. On January 13, the CV^2 started at 0.28. The route destabilized progressively until CV^2 reached 0.5, then returned to more stable conditions, with CV^2 oscillating between 0.15 and 0.35. On January 20, CV^2 started at 0.15 and rapidly increased to 0.6, then stabilized temporarily before starting to rise again. On May 12, during the implementation of the proposed method, the route started with a CV^2 of 0.39, then dipped 45 minutes later for another 45 minutes. Starting at 4 PM, CV^2 increased sharply as the lagging

bus accumulated delay in the peak traffic direction. By 4:45 the route was stabilized and CV^2 was in the same range as it was on January 13 and 20 at the same time.

The holding method was successfully implemented from 2:30 to 5:30 for two days with different results. On Tuesday, May 9, the method was able to dispatch buses with more stable headways than the same day off-peak and than comparable Tuesdays for the same period. On Friday, May 12, the route was severely disrupted right before the implementation, and the proposed method was able to slowly stabilize the route. On both days, the disproportionate impact of small disruptions highlighted the importance of considering dwell time and bathroom breaks explicitly in the holding method. The proposed method was only applied at two control points, instead of the regular four, due to a limited availability of supervisors. Implementing a Hybrid version of the proposed method as in Berrebi et al. (2016) at four control points may substantially reduce the progression of headway variability along the route and even further improve our results.

4.6.3 Georgia Tech

The holding method was continuously implemented from 3:13 to 4:30 on April 21, 2017, on the Georgia Tech Red Stinger Route. Because we were unable to find comparable time periods with valid data, the implementation of the proposed method is presented alone. Figure 25 shows headways, before, during, and after the implementation, as well as holding times throughout. Before the start of the implementation, headways were unstable, fluctuating between 236 and 878 seconds. The implementation almost immediately stabilized the route, maintaining even headways for six consecutive dispatches. At 4:12 PM, one vehicle left the control point with a headway of 646 seconds due to a miscommunication between the dispatcher and the operator. The implementation ended shortly after because one vehicle ran out of gas and had to replenish. The holding times were shorter than 100 seconds before the

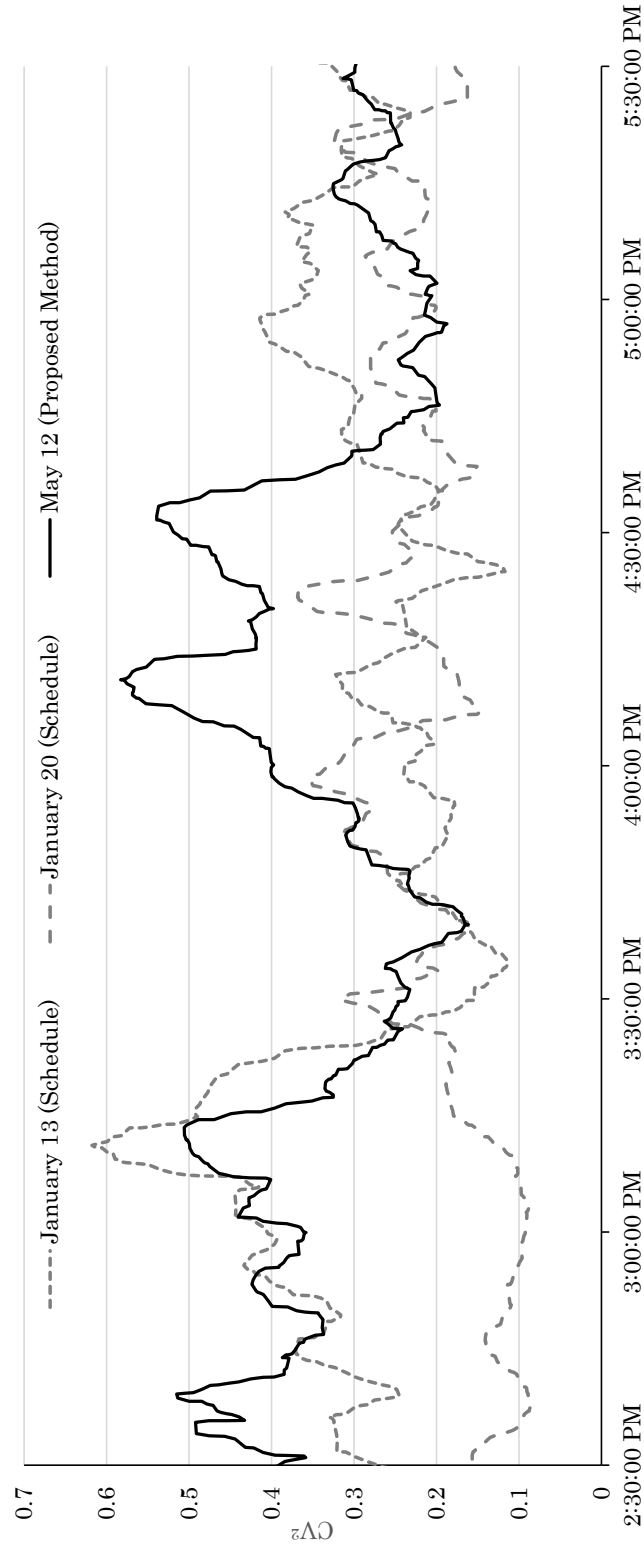


Figure 24: Headway Coefficient of Variation (CV^2) throughout the route between 2:30 and 5:30 under the schedule (January 12 and 20) and under the proposed method (May 12)

implementation because the chosen stop is not the main control point. During the implementation, vehicles were held up to 317 seconds.

4.7 Lessons Learned

In this research, we have found that implementing a real-time holding method involved technical challenges that can be overlooked in a simulation. Transit agencies that wish to apply real-time control should address four elements of the implementation that can affect the performance of the holding method: data, predictions, human element, and surrounding environment. These factors were found to be the greatest source of disruption in all three implementations.

4.7.1 Location Data

The method in Berrebi et al. (2015) uses AVL data in real-time to inform predictions and generate holding times when vehicles arrive at a control point. Simulation experiments, such as Berrebi et al. (2016), assume that perfect vehicle location data are available. In practice, however, AVL data can be unavailable, inaccurate and lagging. This subsection assesses how imperfect data can affect the implementation of the method in Berrebi et al. (2015) or any method.

In most transit systems, vehicle location data are recorded by GPS tracking devices aboard the vehicles and sent to a server via 3G. The GPS unit may be unable to send location data for several minutes at a time due to a defective device, a software problem or a loss of signal. When the tracking device is able to re-establish a connection, it may be unclear how much distance has been traveled. On the Atlanta Streetcar and VIA Route 100 implementations, vehicles were sometimes matched to the route segment traveling in the opposite direction, as if the vehicle had already passed the control point and gone back around. To mitigate this problem, a maximum speed specified how far a vehicle could have realistically traveled during the loss of signal. The maximum speed had to be small enough to avoid matching vehicles to

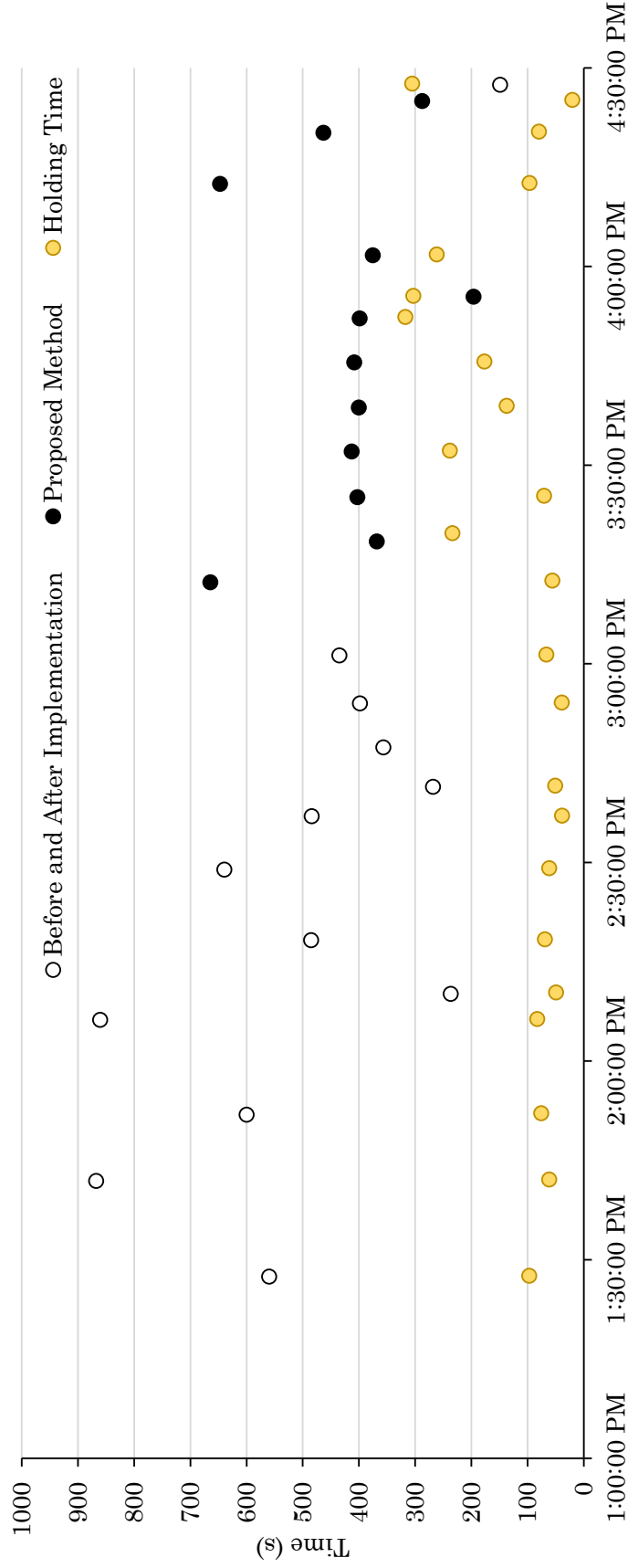


Figure 25: Headway Coefficient of Variation (CV^2) throughout the route between 2:30 and 5:30 under the schedule (January 12 and 20) and under the proposed method (May 12)

the other side of the road and long enough that slow vehicles would still get matched. If not, the vehicle would need to be reassigned manually or have to go all the way around the route before getting matched.

When the connection is lost for more than five minutes, DynamicTime stops considering the predicted arrival of the vehicle at control points. Since the method needs to consider every vehicle on the route, the loss of AVL signal for extended periods of time can undermine the implementation of the holding method. This problem arose frequently on the Georgia Tech Stinger Route, to the point where it was rare that every vehicle on the route was emitting vehicle location data. Although we were able to implement the method for an hour, the AVL data is insufficiently reliable to support a full scale implementation.

The error in vehicle location can differ from one vehicle to the next and fluctuate as vehicles shift from a suburban environment where GPS devices have clear view of the sky to dense urban environment where tall building reverberate the signal in an *urban canyon effect*. In parts of the route that turn, curve, or loop, the wander in GPS signal can cause the system to assign a vehicle to the wrong route segment. To avoid this issue, we specified the maximum matching distance for each segment on the route based on its shape⁴. For example, the maximum matching distance was much shorter at the Northern Control point loop of VIA Route 100, than along Highway I-10.

Vehicle location systems pulse latitude and longitude data at a set frequency into the server. On the Atlanta Streetcar and the Georgia Tech Stinger Route, we were able to access the data from the server directly every 10 and 15 seconds, respectively. On the VIA Route 100, however, the data was deposited on an internal server, then polled through a real-time feed that refreshed every 30 seconds. The additional layer elongated the lag between the time of recording and when the data became available,

⁴<https://github.com/scrudden/core/wiki/Notes>

which took up to 105 seconds. The dashboard did not display holding times until the system had recognized that a vehicle had arrived at the control point, which could lead to confusion for the dispatcher. To mitigate this problem, we started displaying holding times before vehicles arrived at control points based on predicted arrival times. This functionality solved the problem but introduced a new element of complexity that was not needed in the other implementations.

4.7.2 Prediction

The simulation in Berrebi et al. (2016) found that prediction accuracy had almost no effect on either headway stability nor holding time when the standard deviation of error was less than ten percent of the horizon. In all three implementations, the error was almost always within that range. The simple prediction method employed in conjunction with the simplified holding method derived in Lemma 4.4.1 limited the potential for system failure. The 30-second polling frequency of the VIA real-time feed caused no issue in generating the predictions for the simplified version. More granular data would have been necessary to implement the probabilistic version of Berrebi et al. (2015).

Although predicted arrivals were overall satisfactory, vehicles on VIA Route 100 with long headways tended to depart later than recommended from control points. The excess holding time aggravated the already large gaps in service. For example, on May 12, one vehicle on VIA Route 100 arrived at the control point at 15:23, 11 minutes after the last vehicle had left. Although the bus was instructed to depart immediately, it had to spend 3 minutes and 30 seconds boarding and alighting passengers. Unlike most vehicles, which could board and alight passengers while waiting for their recommended departure time, operators who were instructed to depart immediately were substantially delayed. To avoid this issue, future implementation should consider boarding and alighting times into the arrival predictions at the control point

and consider holding time as excess to dwell time.

4.7.3 Human element

In a live implementation, holding instructions must be clearly communicated to vehicle operators to avoid misinterpretation. Misinterpretation of holding instructions happens when the instructions fail to address a perceived and immediate decision. On the Atlanta Streetcar, dispatchers informed operators of their recommended holding time via radio when they arrived at the control point and when it was time to depart. Because operators were unable to gauge the remaining holding time, they often just missed the green phase of the signal, which was located at the control point, and had to wait for an entire cycle. When using a schedule, operators can decide to depart seconds early rather than of minutes late.

Operators have a sense of ground-level operations that may be unperceivable to the holding method, and even to a dispatcher who is away from the field. Indirect lines of communication have the potential to introduce layers of error, which can negatively affect operations. The opacity of the communication process implemented on the Atlanta Streetcar curtailed operators' ability to make judgment calls, which would have helped stabilize the route. When implementing real-time dispatching methods, agencies should provide information to operators and give them some latitude to use their intuition and make decisions

On route 100, VIA considers layovers at control points as recovery time, not breaks. Nonetheless, operators sometimes need to go to the bathroom. In the experiment, operators could use the holding time to go to the bathroom. However, when vehicles were instructed to depart immediately, a bathroom break would delay the departure from the control point. Since the method only ever recommends immediate departures when vehicles are already lagging, bathroom breaks could cause severe disruptions. Since bathroom breaks are unpredictable, the proposed method should

add a short buffer to predicted arrival times to avoid the disruption. Unlike the buffer time in the schedule, this buffer would be added to the predicted time of arrival, and could there be much shorter.

4.7.4 Surrounding environment

The physical environment of a transit route affects both the mean and variance of travel time WSB Parsons Brinckerhoff, Georgia Tech. In the proposed holding method, the travel time for the immediate future is predicted using historical data. These data lose their relevance when the current state of the route deviates from past conditions. When the route characteristics are affected by a discrete event, such as a traffic collision, or when vehicles are re-routed, travel time is also impacted. The performance of the holding method is then affected by its capacity to represent the future based on its knowledge of the past. This, however, is also true of all methods that use historical data to support holding decisions, including the naive schedule. On the Georgia Tech Stinger Route, the implementation period had to be pushed back due to several unexpected construction projects along the route. The route experienced extraordinary delays at times and excessive layovers at other times as the routing kept changing over the course of three weeks.

In order to stabilize headways, vehicles may have to wait for long periods of time at control points. For transit agencies, there may be few eligible locations for control points that can answer two basic conditions. Firstly, control points should be located at stops where many passengers board and alight to avoid delaying passengers who are already on board (Berrebi et al., 2016). Secondly, control points should be in a dedicated space to avoid disrupting surrounding vehicular traffic. The control points used during the implementation were the only stops that fit these criteria.

The northern control point on VIA Route 100 (South Texas Medical Center) does not have enough space for vehicles to pass each other. Therefore, when an operator

stops for a bathroom break, his or her vehicle blocks following buses from departing, which can generate bus bunching from the onset. When possible, transit agencies should select control points where vehicles have enough space to pass each other. An operator going to the bathroom could then switch assignments with the operator of a following vehicle, thereby minimizing the potential for large gaps in service to form. On route 100, however, the option was not available.

Finally, traffic signals are a mediating tool that creates gaps for conflicting traffic. On VIA Route 100, traffic signals can detect buses arriving at intersections. Buses that are more than five minutes late are given priority through green phase extension or red phase truncation. During the implementation, the conditional priority did not help maintain stable headways because vehicles were not adhering to the schedule. In fact, signal timing introduced a new element of randomness that may have deteriorated the quality of predictions. The signals, however, could be modified to give priority to buses based on their headways using the method in Daganzo and Pila-chowski (2011). As shown in (Berrebi et al., 2016), the proposed method can adapt to mid-route control points that apply control based on adjacent headways.

4.8 Conclusion

In this chapter, the holding method recommended in Berrebi et al. (2015) was simplified for the case where predictions are discrete point estimates. The method was successfully implemented on three high-frequency transit routes: the Atlanta Streetcar, the VIA Route 100 in San Antonio, TX, and Georgia Tech Red Stinger Route. The three case-studies are widely distinctive: the first is a streetcar route, the second is a Bus Rapid Transit route, and the third is a campus shuttle. All three run vehicles every 10 minutes or less on a schedule that is unavailable to the public. The three case studies have shown that the proposed method can be implemented on live transit routes to reduce bus bunching.

In each of the three case studies, the proposed method has helped reduce headway variability compared to the schedule that was currently used. On the Atlanta Streetcar, the proposed method was implemented for four three-hour periods. The distribution of headways under the proposed method was more compact than the distribution under the schedule. The proposed method required more holding time and dispatched vehicles with slightly longer headways due to a miscommunication between dispatchers and operators, which led them to miss the permissive phase at a traffic signal located in front of the control point. On VIA Route 100, the holding method was implemented for two three-hour periods. On the first day, the method was able to maintain more stable headways than the schedule, which was implemented on a low traffic day with additional control points. On the second day, the implementation immediately followed an important disruption of the system. The method was able to slowly stabilize the route nonetheless. Finally, on the Georgia Tech shuttle, the method could only be implemented for 75 minutes due to faulty GPS data, and was able to maintain more stable headways than earlier that day. We expect that with the automation the holding protocol and with additional control points throughout the route, these results would further improve.

The three case studies provided insights on the potential challenges in implementing a real-time holding method on high-frequency transit routes. We identified four main factors that need to be addressed to enable a successful implementation: vehicle location data, arrival prediction, human element, and route configuration. The availability, accuracy, and frequency of AVL can have an important impact on the method’s ability to consider and predict vehicles on the route. We found that overall, the quality of predictions generated by TransiTime were satisfactory but that predictions should consider the dwell time at control points to avoid dispatching vehicles before they are ready to depart. It is imperative to provide operators with clear instructions that allow them to understand the holding protocol to make judgment

calls in special circumstances. Finally, the route configuration is important to consider as control point design and traffic signals can have a disproportionate impact on operations.

The institutional frameworks of the three case studies have played an important role in defining the success of the implementation. For the Georgia Tech Parking and Transportation and the City of Atlanta, which both operate small transit systems, the primary motivation for implementing the proposed method was in the research itself. Both agencies were interested in fostering their relationship with our research lab and learn more about their own operations. Due to the size of their system, they had the flexibility to change the dispatching protocol with relative ease. VIA Metropolitan Transit, on the other hand, operates the 24th largest bus system in the United States. The agency had a greater incentive in implementing the proposed method: solving a real problem that costs hundreds of hours of delay for thousands of passengers per day on Route 100 alone. Although the agency has a more rigid framework, it can allocate more resources and take further risks to solve the bus bunching problem.

As academic researchers, we often limit real-world implementations to smaller and closer systems. These systems provide an opportunity to test the research in a safe environment but they rarely face the same scale of problems as larger transit agencies. This research has shown that going into larger and more chaotic fields can put the research to a true test and allow it to solve greater problems. Through this research, we have gained insights on the factors that condition the successful implementation of a real-time holding method on high-frequency routes. The next step is to conduct more sophisticated sensitivity analyses on these factors. The quantity of data required for these tests would require a permanent implementation of the holding method.

CHAPTER V

CONCLUSION

In this thesis, a real-time dispatching method was derived to prevent vehicles from bunching on high-frequency routes. The holding method considers the bus bunching problem globally to diffuse the large gaps caused by lagging vehicles onto the preceding buses. The method was compared to other methods used in practice and recommended in the literature by simulation. A series of sensitivity analyses found that the proposed method could dispatch buses with both the lowest headway variability and the least holding time on a wide range of operating conditions. The method was implemented on three high-frequency transit routes and compared to the schedule currently used. We found that the proposed method outperformed the schedule and identified the factors of success in implementing a real-time dispatching method. Section 5.1 of the conclusion lists the contributions of this thesis and Section 5.2 discusses the implications of the results and the highlights a path research.

5.1 Contributions

5.1.1 A Real-Time Dispatching Policy

- A real-time bus holding method is developed in this thesis. The method can minimize headway variance while still dispatching buses at the rate at which they arrive. It is the first to spread delays evenly amongst buses by considering every vehicle on the route. Unlike holding methods used in practice and recommended in the literature, the proposed approach can yield a route-level natural headway, based on current operating conditions without need for pre-planned operations.

- An analytical solution to the Markovian decision problem to minimize headway variance is derived in this thesis. The optimality properties of solutions are found by backward induction. A close approximation to the optimal decision process is found as a closed form function of the joint probability distribution of bus arrival times. This solution allows to avoid discretizing the state space, in n dimensions, which would be computationally prohibitive.

5.1.2 Simulation of Holding Methods

- The main closed-form holding methods used in practice and recommended in the literature are all compared by a simulation in terms of holding time and headway stability. We explored the trade-offs between holding time and headway stability in different conditions. The sensitivity of each method to its parametrization and to the control point density is evaluated. The results are expressed with dimensionless parameters to extend their generality to any route.
- The sensitivity of prediction-based methods to the prediction quality is evaluated. We established dimensionless measures of both confidence and accuracy and recorded their effect on both holding time and headway stability.
- The method developed in Berrebi et al. (2015) is shown to work well with a prediction method recommended in Hans et al. (2015). The simulation demonstrates for the first time that a prediction method for the probability distribution of each bus arrival time could be sufficiently accurate to yield more stable headways than any other method with the same amount of holding time. The prediction method was also found to perform almost as well as if perfect predictions were available.

5.1.3 Implementation on Three High-Frequency Routes

- The proposed holding method is simplified for the case where predictions are deterministic point estimates. The simplified expression is shown to be equivalent to the original formula in Berrebi et al. (2015).
- The proposed holding method was successfully implemented on three high-frequency transit routes: the Atlanta Streetcar, the VIA Route 100 in San Antonio, TX, and the Georgia Tech Red Stinger Route. In all three systems, the proposed method outperformed the schedule that was currently in place.
- The challenges to implementing a real-time dispatching method are analyzed and discussed. In particular, the importance of vehicle location data quality, prediction accuracy, human element, and route configuration are discussed. This discussion can support the live implementation of the proposed or any other holding method on high-frequency routes.

5.2 *Concluding Remarks and Future Research*

5.2.1 Discussion

In this thesis, the bus bunching problem is addressed as a stochastic decision problem. Holding time is used to create gaps between vehicles that are predicted to arrive at the control point in close succession. The method sets the rate of dispatch as the maximum expected relative frequency of arrival for all vehicles currently on the route. The proposed approach is based on the assumption that at most n vehicles will be dispatched by the time the n^{th} vehicle returns to the control point. Under this constraint, the proposed method minimizes the headway variance and equilibrates the system towards a route-level natural headway.

There is a trade-off between stabilizing headways, which benefits passengers waiting at-stop, and maintaining high-operating speed, which benefits in-vehicle passengers (Furth et al., 2006; Furth and Wilson, 1981; Cats et al., 2011). The primary focus of the proposed method is to minimize the at-stop waiting time of randomly arriving passengers. The proposed method does not explicitly consider the waiting time of passengers who are already boarded in the vehicles. This waiting time can be especially long when one or several vehicles have accumulated substantial delay. This holding time, however is useful to prevent headways from destabilizing even further, which would require even more holding time later.

In Chapter 3, we compared the proposed holding method with other closed-form methods used in practice and recommended in the literature in a series of sensitivity analyses. We tested each method’s trade-off between headway stability and holding time based on its parametrization, the number of control points and the quality of predictions available. We found that the proposed method systematically yields the most stable headways, but requires long holding times (though other methods, including the Naive Headway require even more). The proposed method is therefore adapted for routes where headway dynamics cause severe bus bunching problems. On these routes, the proposed method can stop the progression of headway instability better than any other closed-form method.

On routes where bus bunching frequently occurs, the justification for using the proposed method mainly depends on the control points where it is applied. For passengers, the relative value of headway stability and cost of holding time are only perceived in terms of the delay they experience. The benefit of applying the proposed method is proportional to both the improvement in headway stability over the downstream portion of the route and the number of at-stop passengers it will impact. Conversely, the disbenefit is proportional to both the holding time and the number of passengers who remain boarded at the control point. An important criterion to

justify the application of the method is therefore the existence of stops where few passengers ride through and where upstream stops have great passenger demand.

The criteria for suitable routes and control points is not overly restrictive. In Chapter 4, the proposed method was successfully implemented on three high frequency routes, the Atlanta Streetcar, the VIA Route 100, in San Antonio, TX, and the Red Stinger Route on Georgia Tech Campus. In all three implementations, the proposed method was able to yield more stable operations than the schedule in historical method. The VIA Route 100 case-study, in particular, is prototypical. In historical data, the schedule was unable to prevent buses from bunching between control points. There was limited opportunity for on-route control because the vehicles have no space to hold mid-route. The proposed method was able to stabilize the route, even following a disruption and with fewer control points. These control points are terminal stops, where almost all passengers alight and where new passengers board the vehicle expecting evenly spaced departures. The VIA Route 100 resembles the Tri-Met Route 72, which was used as a model in Chapter 3, and to many routes across the world.

This thesis has developed the first closed-form holding method that can consider every vehicle on the route to minimize headway variance. The sensitivity of the method was compared to other holding methods by simulation and its performance was tested in three live implementations. In order to gain even greater insights in a realistic setting, the holding method would need to be implemented for several consecutive months and compared to similar periods in different years. Since the DynamicTime application is fully open-source, transit agencies facing bus bunching issues can now implement the proposed method independently.

5.2.2 Application to broader transit problems

Public transportation can generate all the benefits of mobility while minimizing the costs of energy consumption, urban sprawl, and pollution. Public transportation makes the most effective use of limited right-of-way by moving a large number of people in a small space. To maximize transit usage and its associated positive externalities, transit agencies strive to provide reliable service by allocating capacity to the demand. Responding early to deficits in capacity can prevent large gaps in service from forming and from spiraling out of control. This task is difficult because centralized transportation systems are inherently unstable. Equilibrium points are stochastic because they shift with their conflicting environment; they are fragile because deficits in capacity tend to accumulate. The current *modus operandi* of transit agencies consists in scheduling operations based on historical data and add buffer in case of disruption. The problem with this approach is that buffer wastes resources when operating conditions are fluid and is insufficient when they are congested.

The research presented in this dissertation is just one example of how transit agencies can take a dynamic approach to a dynamic problem. As transit agencies gain access to an increasing amount and quality of automatically collected data, they become able to make operational decisions in real-time. Using the knowledge of current operating conditions, agencies can predict how the system will evolve in the near future. Decisions based on better predictions can stabilize the system faster while using less buffer time than operations planned months in advance using historical data.

Future research should focus on providing access to high-frequency routes from low density neighborhoods. Transit agencies in the United States are losing ridership, in part due to the declining access to high-capacity service and in part due to competition from demand-responsive mobility providers. In order to remain competitive, transit agencies need to provide better connections to low-density neighborhoods, where fixed

route service is not effective. Vehicles must travel long distances to reach dispersed customers, which spreads the service thin. Declining ridership causes revenue loss which causes further cuts in service. The downward spiral of fixed route transit in low-density neighborhoods leaves a gap in service to connect the first-and-last-mile, which endangers the core network ridership

The approach described in this dissertation could be used to address the first-and-last-mile gap in low-density neighborhoods. Using real-time information about capacity and demand, transit agencies could replace fixed routes by flex routes. Future research should employ stochastic optimization to dispatch demand responsive transit vehicles. A new methodology could route vehicles based on both the current state of the system and its predicted evolution. Simulation studies should evaluate the sensitivity of different dispatching methods to the surrounding environment. Gaining insights on the performance of flex routing methods and potential trade-offs would enable transit agencies to apply the most appropriate method for their service area. One or several methods could then be implemented in a live experiment.

APPENDIX A

This appendix contains the Proofs of Lemmas 2.4.1. and 2.4.2. as well as Corollary 2.4.6 as per Berrebi et al. (2015). The proofs were too long to include in body of Chapter 2. Lemma 2.4.1. says that an optimal policy is always made of optimal sub-policies, dependent only on the last state and decisions. Lemma 2.4.2 demonstrates the criterion of convexity, necessary for the inflexion point approach. Corollary 2.4.6. sets limits to the error of the approximation made in Equations (20) and (22) of Chapter 2. The Lemmas and Corollary are shown below along with their respective proofs.

Lemma 2.4.1.. *Given h_{i-1}^o and $\{A_2, \dots, A_i\}$, for any i , the action h_i^* is defined as follows:*

$$h_i^* = \arg \min_{h_i} \left[(a_i - d_{i-1} + h_i)^2 + \sum_{j=i+1}^n E[(A_j - A_{j-1} + h_j^* - h_{j-1}^*)^2 | A_2, \dots, A_i] \right] \quad (47)$$

At each decision epoch, $\{h_i^, \dots, h_{n-1}^*\}$ minimizes $u_i^\pi(h_{i-1}^o, A_2, \dots, A_i)$.*

Proof. The statement is true for $i = n - 1$ because for any $\{a_2, \dots, a_{n-1}\}$ and h_{n-2}^o , by definition of h_{n-1}^* , we have that for any arbitrary policy $\pi_{n-1} = \{h_{n-1}\}$:

$$(a_{n-1} - d_{n-2} + h_{n-1}^*)^2 + E[(A_n - A_{n-1} - h_{n-1}^*)^2 | A_2, \dots, A_{n-1}] \leq u_{n-1}^\pi(h_{n-2}^o, a_2, \dots, a_{n-1}) \quad (48)$$

Suppose that $\{h_i^*, \dots, h_{n-1}^*\}$ minimizes $u_i^\pi(h_{i-1}^o, a_2, \dots, a_i)$, then given $\{a_2, \dots, a_i\}$ and

h_{i-1}^o we have for any arbitrary policy $\pi = \{h_i^\pi, \dots, h_{n-1}^\pi\}$:

$$(a_i - d_{i-1} + h_i^*)^2 + \sum_{j=i+1}^n E [(A_j - A_{j-1} + h_j^* - h_{j-1}^*)^2 | A_2, \dots, A_i] \quad (49)$$

$$\leq (a_i - d_{i-1} + h_i^\pi)^2 + E \left[(A_{i+1} - A_i + h_{i+1}^* - h_i^\pi)^2 + \sum_{j=i+2}^n (A_j - A_{j-1} + h_j^* - h_{j-1}^\pi)^2 | A_2, \dots, A_i \right] \quad (50)$$

$$\leq (a_i - d_{i-1} + h_i^\pi)^2 + \sum_{j=i+1}^n E [(A_j - A_{j-1} + h_j^\pi - h_{j-1}^\pi)^2 | A_2, \dots, A_i] \quad (51)$$

$$= u_i^\pi(h_{i-1}^o, a_2, \dots, a_i) \quad (52)$$

By backward induction, we have shown that at each decision epoch, the policy

$\{h_i^*, \dots, h_{n-1}^*\}$ minimizes $u_i^*(h_{i-1}^*, A_2, \dots, A_i)$ \square

Lemma 2.4.2.. *The total expected cost criterion is convex in the action space.*

Proof. Suppose that the total expected cost criterion at the first decision epoch is not convex in the action space. Then there exists two action vectors, $\bar{h}' = [h'_1, \dots, h'_{n-1}]^T$ and $\bar{h}'' = [h''_1, \dots, h''_{n-1}]^T$, and a $t \in [0, 1]$ such that:

$$u_1^{t\bar{h}' + (1-t)\bar{h}''} \geq tu_1^{\bar{h}'} + (1-t)u_1^{\bar{h}''} \quad (53)$$

Expanding the total cost criterion gives:

$$\sum_{i=1}^n \int_0^\infty \dots \int_0^\infty f(a_1, \dots, a_n) [-a_{i-1} + a_i + t(-h'_{i-1} + h'_i) + (1-t)(-h''_{i-1} + h''_i)]^2 da_1 \dots da_n \quad (54)$$

$$\geq t \sum_{i=1}^n \int_0^\infty \dots \int_0^\infty f(a_1, \dots, a_n) [-a_{i-1} + a_i - h'_{i-1} + h'_i]^2 da_1 \dots da_n \quad (55)$$

$$\begin{aligned} &+ (1-t) \sum_{i=1}^n \int_0^\infty \dots \int_0^\infty f(a_1, \dots, a_n) [-a_{i-1} + a_i - h''_{i-1} + h''_i]^2 da_1 \dots da_n \\ &= \sum_{i=1}^n \int_0^\infty \dots \int_0^\infty f(a_1, \dots, a_n) [t[-a_{i-1} + a_i - h'_{i-1} + h'_i]^2 + (1-t)[-a_{i-1} + a_i - h''_{i-1} + h''_i]^2] da_1 \dots da_n \end{aligned} \quad (56)$$

Noticing that the k_i term is equal in (54) and (56) for any i and $f(a_1, \dots, a_n)$, we have a contradiction by the Cauchy inequality. As a result there exists no \bar{h}' and \bar{h}'' such that (53) holds. Therefore u_1 is convex and by extension u_i is convex for any i .

□

Lemma 2.4.6.. *If $\epsilon'_j = 0 \forall i$ then Equations (20) and (22) are equivalent to (10) and (11)*

Proof. We will use backward induction to prove Corollary 2.4.6. First note that the hypothesis holds trivially for $i = n - 1$ and $j = n$. Suppose that the hypothesis holds for all $i > i^o$. Then, by the law of iterated expectations, if $\epsilon'_i = 0 \forall i$ we have:

$$\sum_{j=i^o+2}^n E[I_j^{i^o+1}] \left[\frac{E[A_j - A_{i^o+1} | I_j^{i^o+1} = 1, A_2, \dots, A_{i^o+1}]}{j - (i^o + 1)} \right] = E \left[\max_r \frac{A_r - A_{i^o+1}}{r - (i^o + 1)} \right] \quad (57)$$

Then for i^o and $j > i^o$, $E[I_j^{i^o}]$ can be expressed as the sum of probabilities that $I_{i^o+1}^{i^o} = 0$, conditioned on the probabilities of the $E[I_j^{i^o+1}]$ terms as follows:

$$E[I_j^{i^o}] = \sum_{j=i+2}^n P \left(j = \arg \max_r \left[\frac{A_r - A_{i^o+1}}{r - (i^o + 1)} \right] \right) \quad (58)$$

$$\begin{aligned} & P \left(a_{i^o} - d_{i^o-1} \leq \sum_{q=i^o+1}^n E[I_q^{i^o}] \frac{A_q - A_{i^o}}{q - i^o} \mid j = \arg \max_r \left[\frac{A_r - A_{i^o+1}}{r - (i^o + 1)} \right] \right) \\ &= \sum_{j=i+2}^n P \left(j = \arg \max_r \left[\frac{A_r - A_{i^o+1}}{r - (i^o + 1)} \right] \right) \end{aligned} \quad (59)$$

$$\begin{aligned} & P \left(a_{i^o} - d_{i^o-1} \leq \frac{A_j - A_{i^o}}{j - i^o} \mid j = \arg \max_r \left[\frac{A_r - A_{i^o+1}}{r - (i^o + 1)} \right] \right) \\ &= \sum_{j=i+1}^n P \left(j = \arg \max_r \left[\frac{A_r - A_{i^o}}{r - i^o} \right] \right) \end{aligned} \quad (60)$$

We have shown that Corollary 2.4.6. is true for $i = n - 1$ and $j = n$ and that if it is true for some $i^o + 1$, then it is true for i^o . This proves by backward induction that Corollary 2.4.6. is true.

□

Bibliography

- Abkowitz, M., Engelstein, I., 1984. Methods for maintaining transit service regularity. 961.
- Abkowitz, M.D., Lepofsky, M., 1990. Implementing headway-based reliability control on transit routes. *Journal of Transportation Engineering* 116, 49–63.
- Argote-Cabanero, J., Daganzo, C.F., Lynn, J.W., 2015a. Dynamic control of complex transit systems. *Transportation Research Part B: Methodological* 81, 146–160.
- Argote-Cabanero, J., Daganzo, C.F., Lynn, J.W., 2015b. Dynamic control of complex transit systems. *Transportation Research Part B: Methodological* 81, 146–160.
- Atlanta, 2016. Atlanta streetcar system map. URL: <http://streetcar.atlantaga.gov/about/>.
- Barnett, A., 1974. On Controlling Randomness in Transit Operations.
- Bartholdi, J.J., Eisenstein, D.D., 2011. A self-coordinating bus route to resist bus bunching. *Transportation Research Part B: Methodological* .
- Berkow, M., Chee, J., Bertini, R., Monsere, C., 2007. Transit Performance Measurement and Arterial Travel Time Estimation Using Archived AVL Data .
- Berrebi, S., Hans, E., Chiabaut, N., Laval, J., Leclercq, L., Watkins, K., 2016. Comparing holding methods using real-time information. *Transportation Research Part C: Emerging Technologies* (Forthcoming) .
- Berrebi, S., Watkins, K., Laval, J., 2015. A real-time bus dispatching policy to minimize passenger wait on a high frequency route. *Transportation Research Part B: Methodological* doi:doi: 10.1016/j.trb.2015.05.012.

- Boyle, D., Pappas, J., Boyle, P., Nelson, B., Sharfarz, D., Benn, H., 2009. Report 135 Controlling System Costs : Basic and Advanced Scheduling Manuals and Contemporary Issues in Transit Scheduling. Technical Report. The National Academies Press. Washington, D.C.
- Bukkapatnam, S., Dessouky, M., 2003. Distributed architecture for real-time coordination of bus holding in transit networks. *IEEE Transactions on Intelligent Transportation Systems* 4, 43–51. URL: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1200899>, doi:doi: 10.1109/TITS.2003.809769.
- Cathey, F., Dailey, D., 2003. A prescription for transit arrival/departure prediction using automatic vehicle location data. *Transportation Research Part C: Emerging Technologies* 11, 241–264.
- Cats, O., Larijani, A., Koutsopoulos, H., Burghout, W., 2011. Impacts of holding control strategies on transit performance: Bus simulation model analysis. *Transportation Research Record: Journal of the Transportation Research Board* , 51–58.
- Chen, M., Liu, X., Xia, J., 2005. Dynamic Prediction Method with Schedule Recovery Impact for Bus Arrival Time. *Transportation Research Record* 1923, 208–217. doi:doi: 10.3141/1923-22.
- Chien, S.I.J., Ding, Y., Wei, C., 2002. Dynamic bus arrival time prediction with artificial neural networks. *Journal of Transportation Engineering* 128, 429–438.
- Cortés, C.E., Sáez, D., Milla, F., Núñez, A., Riquelme, M., 2010. Hybrid predictive control for real-time optimization of public transport systems’ operations based on evolutionary multi-objective optimization. *Transportation Research Part C: Emerging Technologies* 18, 757–769.

- Daganzo, C.F., 2009. A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons. *Transportation Research Part B: Methodological* 43, 913–921.
- Daganzo, C.F., Pilachowski, J., 2011. Reducing bunching with bus-to-bus cooperation. *Transportation Research Part B: Methodological* 45, 267–277. doi:doi:10.1016/j.trb.2010.06.005.
- Delgado, F., Muñoz, J.C., Giesen, R., Cipriano, A., 2009. Real-Time Control of Buses in a Transit Corridor Based on Vehicle Holding and Boarding Limits. *Transportation Research Record: Journal of the Transportation Research Board* 2090, 59–67.
- Delgado, F., Munoz, J.C., Giesen, R., 2012. How much can holding and/or limiting boarding improve transit performance? *Transportation Research Part B: Methodological* 46, 1202–1217.
- Development, P., 2013. APTA Surveys Transit Agencies on Providing Information and RealTime Arrivals to Customers , 1–13.
- Ding, Y., 2000. The Development and Application of Dynamic Models for Predicting Transit Arrival Times. Technical Report.
- Eberlein, X.J., Wilson, N.H., Bernstein, D., 2001. The holding problem with real-time information available. *Transportation science* 35, 1–18.
- Fan, W.D., Machemehl, R.B., 2009. Do transit users just wait for buses or wait with strategies? *Transportation Research Record: Journal of the Transportation Research Board* 2111, 169–176.

- Furth, P., Hemily, B., Muller, T., Strathman, J., 2006. TCRP Report 113: Using Archived AVL-APC Data to Improve Transit Performance and Management. Technical Report. The National Academies Press. Washington, D.C.
- Furth, P.G., Muller, T.H.J., 2007. Service Reliability and Optimal Running Time Schedules. *Transportation Research Record* 2034, 55–61. doi:doi: 10.3141/2034-07.
- Furth, P.G., Wilson, N.H., 1981. Setting frequencies on bus routes: theory and practice. *Transportation Research Record* 818, 1–7.
- Gauss, C.F., 1809. *Theoria motus corporum coelestium in sectionibus conicis solem ambientium auctore Carolo Friderico Gauss. sumtibus Frid. Perthes et IH Besser.*
- Grisby, D., 2013. APTA Surveys Transit Agencies on Providing Information and RealTime Arrivals to Customers , 1–13.
- Gurmu, Z.K., Fan, W.D., 2014. Dynamic travel time prediction models for buses using only gps data, in: *Transportation Research Board 93rd Annual Meeting.*
- Hans, E., Chiabaut, N., Leclercq, L., 2014. Investigating the irregularities of bus routes. *Journal of Advanced Transportation* 45, 117–128. doi:doi: 10.1002/atr.
- Hans, E., Chiabaut, N., Leclercq, L., Bertini, R., 2015. Real-time bus route state forecasting using particle filter. *Transportation Research Part C: Emerging Technologies* .
- Hickman, M.D., 2001. An Analytic Stochastic Model for the Transit Vehicle Holding Problem. *Transportation Science* 35, 215–237.
- Jeong, R., Rilett, L., 2005. Prediction Model of Bus Arrival Time for Real-Time Applications. *Transportation Research Record* 1927, 195–204. doi:doi: 10.3141/1927-23.

- Jeong, R., Rilett, L.R., 2004a. Artificial Neural Network Model. IEEE Intelligent Transportation Systems Conference , 988–993.
- Jeong, R., Rilett, L.R., 2004b. Bus arrival time prediction using artificial neural network model, in: Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on, IEEE. pp. 988–993.
- KFH, 2013. Transit capacity and quality of service manual.
- Kittelson, 2003. Transit capacity and quality of service manual. volume 100. Transportation Research Board.
- Levinson, H.S., 1991. Supervision strategies for improved reliability of bus routes. 15.
- Lizana, P., Muñoz, J.C., Giesen, R., Delgado, F., 2014. Bus control strategy application: case study of santiago transit system. Procedia Computer Science 32, 397–404.
- Luethi, M., Weidmann, U., Nash, A., Weidmann, U., Weidmann, U., 2006. Passenger arrival rates at public transport stations. Institute for Transport Planning and Systems, ETH Zurich.
- Mazloumi, E., Rose, G., Currie, G., Moridpour, S., 2011. Prediction intervals to account for uncertainties in neural network predictions: Methodology and application in bus travel time prediction. Engineering Applications of Artificial Intelligence 24, 534–542. doi:doi: 10.1016/j.engappai.2010.11.004.
- Milkovits, M.N., 2008. Simulating Service Reliability of a High Frequency Bus Route Using Automatically Collected Data by Master of Science in Transportation .
- Newell, F., Potts, R.B., 1964. Maintaining a bus schedule. volume 2.
- Newell, G., 1974. Control of Pairing Vehicles on a public Transportation Route, Two Vehicles, One Control Point .

- Oort, N.V., Wilson, N.H.M., Nes, R.V., 2009. Reliability improvement in short headway transit services : schedule-based and headway-based holding strategies. *Transportation Research Board* , 1–19.
- Osuana, E., Newell, G., 1972. Control Strategies for an Idealized Public Transportation System , 52–72.
- Pangilinan, C., Wilson, N., Moore, A., 2008. Bus supervision deployment strategies and use of real-time automatic vehicle location for improved bus service reliability. *Transportation Research Record: Journal of the Transportation Research Board* , 28–33.
- Parking, G.T., Services, T., 2017. Via primo map. URL: <http://pts.gatech.edu/sites/default/files/documents/stinger-buses-map.pdf>.
- Program, T.C.R., 2013. Transit Capacity and Quality of Service Manual 3rd Edition. Technical Report. Transportation Research Board.
- Puterman, M.L., 2009. Markov decision processes: discrete stochastic dynamic programming. volume 414. Wiley. com.
- Sánchez-Martínez, G., Koutsopoulos, H., Wilson, N., 2016. Real-time holding control for high-frequency transit with dynamics. *Transportation Research Part B: Methodological* .
- Sánchez-Martínez, G.E., Koutsopoulos, H.N., Wilson, N.H., 2015. Event-driven holding control for high-frequency transit. *Transportation Research Record: Journal of the Transportation Research Board* , 65–72.
- Shalaby, A., Farhan, A., 2004. Prediction model of bus arrival and departure times using avl and apc data. *Journal of Public Transportation* 7, 41–62.

- Strathman, J.G., Kimpel, T.J., Kenneth, J., Gerhart, R.L., Callas, S., 2002. Evaluation of transit operations : data applications of Tri-Met's automated Bus Dispatching System , 321–345.
- Sun, D., Luo, H., Fu, L., Liu, W., Liao, X., Zhao, M., 2007. Predicting Bus Arrival Time on the Basis of Global Positioning System Data. *Transportation Research Record* 2034, 62–72.
- TriMet, 2016. Line 72 route map. URL: <http://trimet.org/schedules/r072.htm#map>.
- Turnquist, M.A., 1982. Strategies for improving bus transit service reliability. Technical Report.
- Van Oort, N., Wilson, N.H., Van Nes, R., 2010. Reliability improvement in short headway transit services. *Transportation Research Record: Journal of the Transportation Research Board* 2143, 67–76.
- VIA, 2017. Via primo map. URL: <http://www.viaprime.com/BusService/Schedules.aspx>.
- Walker, J., 2011. *Human Transit: How Clearer Thinking about Public Transit Can Enrich Our Communities and Our Lives*. Island Press.
- Winston, W.L., Goldberg, J.B., 2004. *Operations research: applications and algorithms*. Thomson/Brooks/Cole Belmont.
- WSB Parsons Brinckerhoff, Georgia Tech, .
- Xuan, Y., Argote, J., Daganzo, C.F., 2011. Dynamic bus holding strategies for schedule reliability: Optimal linear control and performance analysis. *Transportation Research Part B: Methodological* 45, 1831–1845.

- Zhao, J., Dessouky, M., Bukkapatnam, S., 2006. Optimal Slack Time for Schedule-Based Transit Operations. *Transportation Science* 40, 529–539. URL: <http://transci.journal.informs.org/cgi/doi/10.1287/trsc.1060.0170>, doi:doi: 10.1287/trsc.1060.0170.
- Zolfaghari, S., Azizi, N., Jaber, M.Y., 2004. A model for holding strategy in public transit systems with real-time information. *International Journal of Transport Management* 2, 99–110.